



# Probing patchy reionization through tau-21cm correlation statistics

## Citation

Meerburg, P. Daniel, Cora Dvorkin, and David N. Spergel. 2013. Probing patchy reionization through tau-21cm correlation statistics. *The Astrophysical Journal* 779, no. 2: 124.  
doi:10.1088/0004-637x/779/2/124.

## Published Version

doi:10.1088/0004-637x/779/2/124

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:29995348>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# PROBING PATCHY REIONIZATION THROUGH $\tau$ -21CM CORRELATION STATISTICS

P. DANIEL MEERBURG<sup>1</sup>, CORA DVORKIN<sup>2</sup>, AND DAVID N. SPERGER<sup>1</sup>

<sup>1</sup>Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08540 USA. and

<sup>2</sup>Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA

(Dated: October 17, 2013)  
 Draft version October 17, 2013

## ABSTRACT

We consider the cross-correlation between free electrons and neutral hydrogen during the epoch of reionization. The free electrons are traced by the optical depth to reionization  $\tau$  while the neutral hydrogen can be observed through 21 cm photon emission. As expected, this correlation is sensitive to the detailed physics of reionization. Foremost, if reionization occurs through the merger of relatively large halos hosting an ionizing source, the free electrons and neutral hydrogen are anti-correlated for most of the reionization history. A positive contribution to the correlation can occur when the halos that can form an ionizing source are small. A measurement of this sign change in the cross-correlation could help disentangle the bias and the ionization history. We estimate the signal-to-noise of the cross-correlation using the estimator for inhomogeneous reionization  $\hat{\tau}_{\ell m}$  proposed by Dvorkin and Smith (2009). We find that with upcoming radio interferometers and CMB experiments, the cross-correlation is measurable going up to multipoles  $\ell \sim 1000$ . We also derive parameter constraints and conclude that, despite the foregrounds, the cross-correlation proofs a complementary measurement of the EoR parameters to the 21 cm and CMB polarization auto-correlations expected to be observed in the coming decade.

**Keywords:** Reionization, 21 centimeters, Cosmic Microwave Background

## 1. INTRODUCTION

The epoch of reionization (EoR) is one of the least understood periods of cosmic history, with only limited observational measurements (see e.g. Loeb and Barkana (2001)). The absence of the Gunn-Peterson trough in the spectra of quasars implies that reionization should have been completed around  $z = 6$  (Fan et al. 2006a,b). The total optical depth to reionization has been measured to be  $\tau = 0.084 \pm 0.013$  (Hinshaw et al. 2012). If reionization is assumed to be instantaneous, this would imply a transition redshift of  $z_{\text{re}} = 11$ . Besides these constraints, we know very little about the details of reionization, such as the typical halo mass associated with the first ionizing objects as well as the distribution of these objects inside the halos.

The spectral mapping of neutral hydrogen in emission (Hogan and Rees 1979; Scott and Rees 1990; McQuinn et al. 2006; Madau et al. 1997; Zaldarriaga et al. 2004; Furlanetto et al. 2004a,b, 2006) promises to be a new probe of the EoR. The spontaneous hyperfine spin flip transition causes the emission of a photon with a wavelength of 21 centimeters in the rest frame. Applying different base filters to the observed emission, it is possible to map the distribution of neutral hydrogen in the Universe as a function of redshift. The auto-correlation of the observed maps is very sensitive to the EoR parameters.

Additionally, cross correlating the observed maps with other observables could provide complementary constraints on the EoR parameters. For example, the 21 cm fluctuations are expected to be correlated with galaxies (Lidz et al. 2009; Wiersma et al. 2012) as well as

with the Cosmic Microwave Background (CMB) fluctuations through the Doppler peak and the kinetic SZ effect (see e.g. Alvarez et al. (2006); Salvaterra et al. (2005); Adshead and Furlanetto (2007); Tashiro et al. (2008); Tashiro et al. (2010, 2011); Holder et al. (2006) and (Jelić et al. 2010; Natarajan et al. 2012) for recent simulations). Unlike the fluctuations in 21 cm, which are a direct representation of the underlying neutral hydrogen, no direct measurement of the electron density at high redshifts exists. The electron density can be measured indirectly through its integrated effect on the CMB, providing us with a number ( $\tau$ ) that tells us the fraction of photons affected by scattering of electrons along the line of sight. One can go further and reconstruct the inhomogeneities in the optical depth field by considering second order effects on the CMB due to the screening mechanism (Dvorkin et al. 2009), Thomson scattering and the kSZ effect (see Dvorkin and Smith (2009), where an estimator of the anisotropic optical depth field is derived and Gluscevic et al. (2012) for an implementation of this estimator to WMAP 7-year data). In this work we consider the cross-correlation between a reconstructed map of the inhomogeneous optical depth  $\tau_{\ell m}$  (using CMB polarization observations) and a neutral hydrogen map measured through the redshifted 21 cm lines. Intuitively, these two observables are expected to be anti-correlated on most scales. We would like to stress that this cross-correlation is fundamentally different from direct cross correlations between CMB temperature and polarization and 21 cm maps (Tashiro et al. 2008; Tashiro et al. 2010); the  $\tau$  estimator is a quadratic estimator, making this cross correlation a statistical 3 point correlation function rather than a 2 point function.

Besides providing complementary constraints on the EoR parameters, the cross-correlation between the 21

cm field and the CMB should in principle be less sensitive to the details of the foregrounds. Although current (Masui et al. 2012; Chapman et al. 2012) and upcoming experiments (Harker et al. 2010; Mellema et al. 2012) are expected to be capable of measuring the auto-correlation of 21 cm maps, one very persistent nuisance in extracting the signal from reionization are the foregrounds. Inhomogeneities in the 21 cm signal due to patchy reionization must be separated from fluctuations in foreground sources. Typical foreground sources are faint radio galaxies, starburst galaxies and galaxies responsible for reionization. In addition, our own Galaxy is very bright at the frequencies one aims at for mapping the 21 cm signal from reionization, exceeding the 21 cm reionization signal by several orders of magnitude. Attempts have been made to characterize these foregrounds (Liu et al. 2009; Jelić et al. 2008; Bernardi et al. 2010, 2009; Liu et al. 2012). Despite these efforts, foregrounds can never be fully removed, simply because we do not know their exact origin.

This paper is organized as follows. We review the physics of reionization and derive the expressions for the fluctuations in 21 cm  $a_{\ell m}^{21}$  and fluctuations in the optical depth  $\tau_{\ell m}$  in §2. The former is proportional to the neutral hydrogen fraction, while the latter is proportional to the free electron fraction. Using a simple reionization model (Furlanetto et al. 2004b; Wang and Hu 2006) where regions of HII are represented by spherical bubbles of typical size  $\bar{R}$  we derive an expression for the cross-correlation  $\langle \tau_{\ell m} a_{\ell m}^{21*} \rangle$  in section §3. In §5 and §6 we study the one-bubble and two-bubble contributions to the power spectrum. We compute the angular power spectrum of the cross-correlation in §7. We assess the observability of the cross correlation by using a redshift weighting to maximize the signal to noise. We end this section with an estimate of the EoR parameter constraints, when considering LOFAR and SKA noise levels. We present our conclusions in §9. In the Appendix we discuss the dependence of the cross-correlation on the parameters of the reionization model.

Unless specifically mentioned, we use the following set of parameter values throughout the paper:  $h = 0.704$ ,  $\Omega_b = 0.044$ ,  $\Omega_c = 0.23$ ,  $\Omega_K = 0$ ,  $n_s = 0.96$  and  $\tau = 0.084$ . Lensing is included and we use the non-linear halo fit model to determine the power spectrum of density fluctuations. We use the WMAP pivot scale  $k_* = 0.002 \text{ Mpc}^{-1}$  and  $A_s = 2.46 \times 10^{-9}$  (Hinshaw et al. 2012).

## 2. 21 CM BRIGHTNESS AND THE OPTICAL DEPTH TO REIONIZATION

In this section we will review the standard results for fluctuations in the 21 cm temperature brightness and relate those to fluctuations in the neutral hydrogen fraction. In the second half of this section, we will derive the fluctuations in the optical depth  $\tau$  caused by fluctuations in the free electron fraction along the line of sight, confirming the results first obtained in (Holder et al. 2006).

The optical depth of a region of the IGM in the hyperfine transition is given by (Field 1958)

$$\tau_{21}(z) = \frac{3c^3 h A_{10}}{32\pi k \nu_0^2 T_S} \frac{n_{HI}}{(1+z)(dv_{\parallel}/dr_{\parallel})}, \quad (1)$$

where  $\nu_0 = 1420.4 \text{ MHz}$  is the rest frame hyperfine 21 cm

( $\nu_0 = \lambda_{21}/c$ ) transition frequency,  $A_{10} = 2.85 \times 10^{-15} \text{ s}^{-1}$  is the spontaneous emission coefficient for this transition,  $T_S$  is the spin temperature of the IGM, weighting the relative population of the atoms in the singlet state to atoms in the triplet state (Field 1958),  $n_{HI}$  is the neutral hydrogen density, and  $v_{\parallel}$  the proper velocity along the line of sight. At high redshifts, where peculiar motions along the line of sight are small compared to the Hubble flow,  $dv_{\parallel}/dr_{\parallel} = H(z)/(1+z)$ . At  $z = 10$  dark energy and radiation are both unimportant and we can solve for  $H(z)$  in a matter dominated Universe,  $H(z) \simeq H_0 \Omega_m^{1/2} (1+z)^{3/2}$ .

We can now write the following expression for the optical depth:

$$\tau_{21}(z) \simeq 8.6 \times 10^{-3} (1 + \delta_b) x_H \left[ \frac{T_{cmb}}{T_S} \right] \left[ \frac{1 - Y_p}{1 - 0.248} \right] \times \left( \frac{\Omega_b}{0.044} \right) \left[ \left( \frac{0.27}{\Omega_m} \right) \left( \frac{1+z}{10} \right) \right]^{1/2} \quad (2)$$

Here we used  $T_{cmb} = 2.73(1+z) \text{ K}$ ,  $\delta_b = (\rho_b - \bar{\rho}_b)/\bar{\rho}_b$  and

$$n_{HI} \simeq (1 - Y_p) x_H \frac{\Omega_b}{\Omega_m} \frac{\rho_m}{m_p},$$

where  $x_H$  is the neutral hydrogen fraction, i.e.  $x_H = n_{HI}/(n_{HI} + n_e)$ ,  $\rho_m$  is the matter energy density and  $m_p$  is the proton mass. The factor  $(1 - Y_p)$  addresses the fact that not all protons are in hydrogen, but a fraction is in Helium.

The intensity along the line of sight from a thermal source is given by

$$I = I_0 e^{-\tau} + \int_{\tau}^0 d\tau' e^{-\tau'} \frac{\eta_{\nu}}{\kappa_{\nu}}, \quad (3)$$

with  $\kappa_{\nu}$  the absorption coefficient and  $\eta_{\nu}$  the emissivity of photons. Using  $dI = \eta_{\nu} dl - \kappa_{\nu} I dl = 0$  in the Rayleigh-Jeans limit, we have  $I = 2kT_b \nu^2/c^2$ , while  $\eta_{\nu}/\kappa_{\nu} = 2kT_S \nu^2/c^2$  and  $I_0 = 2kT_{cmb} \nu^2/c^2$ . Hence, we can write the 21 cm brightness temperature as:

$$T_b = T_{cmb} e^{-\tau_{21}} + T_S (1 - e^{-\tau_{21}}) \quad (4)$$

The brightness temperature increment is defined at an observed frequency  $\nu$  corresponding to a redshift  $1+z = \nu_0/\nu$  as

$$\delta T_b(z) \equiv \frac{T_b - T_{cmb}}{1+z} \simeq \frac{(T_S - T_{cmb})}{1+z} \tau_{21} \quad (5)$$

Using Eq. (2), we can re-write Eq. (5) as (Scott and Rees 1990; Madau et al. 1997)

$$\delta T_b(z) \simeq 27 \text{ mK} (1 + \delta_b) x_H \left[ \frac{T_S - T_{cmb}}{T_S} \right] \left[ \frac{1 - Y_p}{1 - 0.248} \right] \times \left( \frac{\Omega_b}{0.044} \right) \left[ \left( \frac{0.27}{\Omega_m} \right) \left( \frac{1+z}{10} \right) \right]^{1/2} \quad (6)$$

There are usually two types of filters associated with the resolution of the experiment. First, there is a finite angular resolution, which will affect all modes perpendicular to the line of sight. Second, since the brightness temperature of the 21 cm emission is a 3-dimensional

field, we are confined to a frequency resolution or bandwidth, which affects the modes along the line of sight (or, equivalently, a redshift resolution).

The total integrated 21 cm surface brightness is given by

$$T_b(\hat{n}, \chi) = T_0(\chi) \int d\chi' W_\chi(\chi') \psi(\hat{n}, \chi') \quad (7)$$

Here  $W_\chi$  is an experimental band filter that is due to the finite frequency resolution of the instrument, which is centered around  $\chi$  (comoving distance). We define the dimensionless brightness temperature  $\psi$  as

$$\psi = (1 + \delta_b) x_H \left( \frac{T_S - T_{cmb}}{T_S} \right) \quad (8)$$

In the limit of  $T_S \gg T_{cmb}$ ,  $\psi = (1 + \delta_b) x_H$ .

Now  $T_0(z)$  can be written as

$$T_0(z) \simeq 27 \text{ mK} \left[ \frac{1 - Y_p}{1 - 0.248} \right] \left( \frac{\Omega_b}{0.044} \right) \left[ \left( \frac{0.27}{\Omega_m} \right) \left( \frac{1+z}{10} \right) \right]^{1/2} \quad (9)$$

We will now consider fluctuations in the free electron density, which in turn will induce fluctuations in the optical depth. The optical depth to distance  $\chi$  along the line of sight is given by

$$\tau(\hat{n}, \chi) = \sigma_T \int_0^\chi d\chi' n_e(\hat{n}, \chi') a(\chi'), \quad (10)$$

where  $\sigma_T$  is the Thomson cross-section,  $n_e$  is the electron number density and  $a$  is the scale factor. Relating the free electron density to the free electron fraction  $x_e$ , we can write

$$n_e(\hat{n}, \chi) \simeq \frac{x_e \rho_b}{m_p} \left( 1 - \frac{3}{4} Y_p \right), \quad (11)$$

assuming that Helium is singly ionized.

The average baryon density diffuses in an expanding background as  $a^{-3}$ , and the free electron density becomes

$$n_e(\hat{n}, \chi) = \left( 1 - \frac{3}{4} Y_p \right) \frac{\rho_{b,0}}{m_p} a^{-3} (1 + \delta_b) x_e \quad (12)$$

The optical depth can in turn be written as

$$\begin{aligned} \tau(\hat{n}, \chi) &= \sigma_T \left( 1 - \frac{3}{4} Y_p \right) \frac{\rho_{b,0}}{m_p} \int_0^\chi \frac{d\chi'}{a^2(\chi')} x_e(\hat{n}, \chi') \\ &\times (1 + \delta_b(\hat{n}, \chi')) \end{aligned} \quad (13)$$

Therefore, we can relate fluctuations in the optical depth  $\delta\tau$  to fluctuations in the 21 cm brightness temperature  $\delta T_b$  (Holder et al. 2006) as

$$\delta\tau = \left( 1 - \frac{3}{4} Y_p \right) \frac{\sigma_T \rho_{b,0}}{m_p H_0 \Omega_m^{-1/2}} \int dz \left[ (1+z)^{1/2} \delta_b - \frac{\delta T_b(z)}{8.5 \text{ mK}} \right], \quad (14)$$

where we have assumed a delta window function.

It is worth noticing that the above expression for the fluctuations in the 21 cm brightness temperature is only

valid for  $T_s > T_{\text{CMB}}$ . Early on, when the number of ionizing sources are rare and the temperature of the IGM close to these sources is coupled to the kinetic temperature by Ly $\alpha$  photons associated with these local sources, this assumption breaks down, and the 21 cm signal can appear in absorption. We neglect this effect in this paper.

### 3. CORRELATING $X$ AND $\psi$

We will now cross-correlate the optical depth fluctuations with the temperature brightness. As we saw before, the CMB optical depth is proportional to the free electron density. If reionization is inhomogeneous, the free electron density is a function of position in the sky. Anisotropies in the optical depth produce three effects in the CMB: (i) screening of the temperature and polarization fluctuations that we observe today by an overall factor of  $e^{-\tau(\hat{n})}$ . This effect generates CMB B-mode polarization; (ii) Thomson scattering: new polarization is generated by scattering of the local temperature quadrupole that each electron sees along the line of sight. This effect also produces B-modes; and (iii) new temperature anisotropy is generated from the radial motion of ionized bubbles relative to the observer (the kinetic Sunyaev Zel'dovich effect).

The two-point correlation function between the E-modes and the B-modes generated from patchy reionization is proportional to the anisotropic part of the optical depth. This fact allowed the authors in Ref. (Dvorkin and Smith 2009) to write a minimum variance quadratic estimator  $\hat{\tau}_{\ell m}$  for the field  $\tau(\hat{n})$ . In this work, we will use the CMB polarization fluctuations to reconstruct a map of  $\tau$  and cross-correlate it with the 21 cm field.

We will use the shorthand notation  $X(\hat{n}, \chi) = x_e(1 + \delta_b)$ . We can write the cross-correlation between the field  $X$  (measured through the CMB) and the field  $\psi$  (measured through 21 cm) as

$$\begin{aligned} \xi_{X\psi} &\simeq -\xi_{xx}(1 + \xi_{\delta\delta}) - (\bar{x}_H - \bar{x}_e + \xi_{x\delta})\xi_{x\delta} \\ &+ (\bar{x}_H - \bar{x}_H^2)\xi_{\delta\delta} \end{aligned} \quad (15)$$

Here we defined  $\xi_{xx} = \langle x_H(\vec{x}_1) x_H(\vec{x}_2) \rangle - \bar{x}_H^2$ ,  $\xi_{\delta\delta} = \langle \delta_b(\vec{x}_1) \delta_b(\vec{x}_2) \rangle$  and  $\xi_{x\delta} = \langle \delta_b(\vec{x}_1) x_H(\vec{x}_2) \rangle$ . We make the simplistic assumption that the connected part of  $\langle \delta_x \delta_x \delta_b \delta_b \rangle$  vanishes. Here  $\delta_x$  corresponds to fluctuations in the neutral hydrogen fraction, which is given by  $x_H = \bar{x}_H (1 + \delta_x)$ .

Before we can compute  $\xi_{X\psi}$  we need to specify our model of reionization. We will assume that the Universe reionized through the growth of ionized bubbles associated with massive halos. The bubbles themselves contain a single source and we assume their size to be larger than the non-linear scale.

We will adopt the following average reionization fraction as a function of redshift

$$\bar{x}_e(z) = \frac{1}{2} \left[ 1 + \tanh \left( \frac{y_{\text{re}} - (1+z)^{3/2}}{\Delta y} \right) \right], \quad (16)$$

which is the one used in the code CAMB (Lewis et al. 2000). Here  $y(z) = (1+z)^{3/2}$ ,  $y_{\text{re}} = y(z_{\text{re}})$  and  $\Delta y$  are free parameters that satisfy our integrated optical depth along the line of sight  $\tau = 0.084$ .

The ionized bubble around a given source is assumed to be spherical with an average radius  $\bar{R}$ . We will assume that the typical ionized bubble radii are log-normal distributed (Zahn et al. 2006), i.e. there is a skewness towards smaller bubble sizes,

$$P(R) = \frac{1}{\bar{R}} \frac{1}{\sqrt{2\pi\sigma_{\ln R}^2}} e^{-[\ln(R/\bar{R})]^2/(2\sigma_{\ln R}^2)}, \quad (17)$$

where  $\sigma_{\ln R}$  is the variance of the distribution.

The average bubble volume is then given by

$$\langle V_b \rangle = \int dR P(R) V_b(R) = \frac{4\pi\bar{R}^3}{3} e^{9\sigma_{\ln R}^2/2} \quad (18)$$

Hence, we can define a volume weighted radius,  $R_0$  such that  $\langle V_b \rangle = 4\pi R_0^3/3$ , which can be written as

$$R_0 = \bar{R} e^{3\sigma_{\ln R}^2/2} \quad (19)$$

If we assume that a given point in space is ionized with Poisson probability, we can write the ionization fraction as

$$\langle x_e(\vec{x}) \rangle_P = 1 - e^{-n_b(\vec{x}) \langle V_b \rangle}, \quad (20)$$

with  $n_b$  the number density of bubbles. The brackets around  $x_e$  are placed to remind us that we are considering a Poisson distribution of sources, and the result is averaged over the Poisson process. We further assume that the number density of bubbles traces the large-scale structure with some bias  $b$ :

$$n_b(\vec{x}) = \bar{n}_b (1 + b\delta_W(\vec{x})), \quad (21)$$

while the average bubble number density is related to the mean ionization fraction as

$$\bar{n}_b = -\frac{1}{\langle V_b \rangle} \ln(1 - \bar{x}_e) \quad (22)$$

Here  $\delta_W$  is the matter over-density  $\delta$  smoothed by a top hat window of radius  $R$ ,

$$\delta_W(\vec{x}) = \int d^3x' \delta(\vec{x}') W_R(\vec{x} - \vec{x}') \quad (23)$$

In momentum space,

$$W_R(k) = \frac{3}{(kR)^3} [\sin(kR) - kR \cos(kR)], \quad (24)$$

which is the Fourier transform of  $W_R(x) = V_b^{-1}$  for  $x \leq R$  and  $W_R(x) = 0$  otherwise.

We further define

$$\langle W_R \rangle(k) = \frac{1}{\langle V_b \rangle} \int_0^\infty dR P(R) V_b(R) W_R(kR) \quad (25)$$

and

$$\langle W_R^2 \rangle(k) = \frac{1}{\langle V_b \rangle^2} \int_0^\infty dR P(R) [V_b(R) W_R(kR)]^2 \quad (26)$$

#### 4. REIONIZATION PARAMETERS

Reionization can only proceed if the seed halo is massive enough for cooling. In particular, line cooling and

atomic cooling are important for collapse. One can relate the virial temperature of the halo to the mass of the halo

$$\frac{T_{\text{vir}}}{10^4 \text{K}} = 1.1 \left( \frac{\Omega_m h^2}{0.15} \right)^{1/3} \left( \frac{1+z}{10} \right) \left( \frac{M}{10^8 M_\odot} \right)^{2/3} \quad (27)$$

Therefore, setting a condition on the amount of cooling necessary to form ionizing objects sets a typical mass of the halo, which will be redshift dependent. We can relate the mean number density of bubbles, the average reionization fraction and the typical bubble volume by inverting Eq. (22):

$$\langle V_b \rangle = -\frac{1}{\bar{n}_b} \ln(1 - \bar{x}_e) \quad (28)$$

The mean bubble number density introduced in Eq. (22) is derived through an integral over the halo mass function, with a mass threshold  $M_{\text{th}}$  which can roughly be set by the virial temperature in Eq. (27)<sup>1</sup>:

$$\bar{n}_b = \int_{M_{\text{th}}}^\infty \frac{dn_h}{d \ln M} \frac{dM}{M} \quad (29)$$

We use the Sheth and Tormen (Sheth and Tormen 2002) halo mass function

$$\frac{dn_h}{d \ln M} = \frac{\rho_{m(0)}}{M} f(\nu) \frac{d\nu}{d \ln M} \quad (30)$$

with

$$\nu f(\nu) = A \sqrt{\frac{2}{\pi}} a \nu^2 (1 + (a\nu^2)^{-p}) e^{-a\nu^2/2} \quad (31)$$

where  $\nu = \delta_c / \sigma_{\text{lin}}(M, z)$  and  $\sigma_{\text{lin}}$  is the variance of the density field smoothed with the top-hat window function enclosing a mass  $M$ :

$$\sigma_{\text{lin}}^2(M, z) = \int \frac{dk}{k} \Delta_m^2(k, z) W_{R(M)}^2(k) \quad (32)$$

It is straightforward to show that

$$\frac{d\nu}{d \ln M} = -\nu \frac{d \ln \sigma_{\text{lin}}(M, z)}{d \ln M} \quad (33)$$

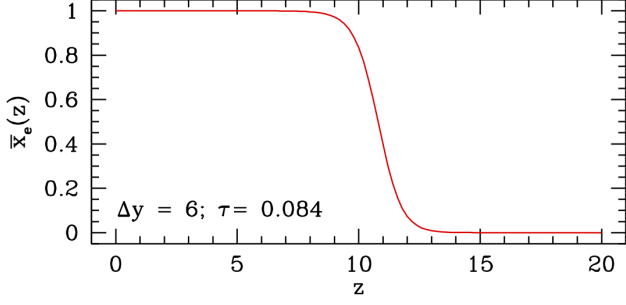
The parameters  $\delta_c$ ,  $a$  and  $p$  can be fitted from simulations.  $A$  is then derived through the constraint  $\int d\nu f(\nu) = 1$ . Consequently, by setting  $M_{\text{th}}$  we can find an expression for the average bubble volume, and we can infer a function of the average bubble radius as a function of redshift (in the assumption of a log normal distribution of radii at any given redshift).

Likewise, the bubble bias can be related to the halo bias (see e.g. Wang and Hu (2006)) as:

$$b = \frac{1}{\bar{n}_b} \int_{M_{\text{th}}}^\infty b_h(M) \frac{dn_h}{d \ln M} \frac{dM}{M} \quad (34)$$

<sup>1</sup> Although the model we are using here is self consistent, relating our toy-reionization model to all other relevant parameters, we find that the resulting bubble radius as a function of redshift is too small compared to simulations (Shin et al. 2008). One can alleviate this discrepancy somewhat by raising  $\bar{M}_{\text{th}}$ . For that purpose we assume the critical temperature to form an ionizing object to be five times the virial temperature.





**Figure 1.** The average ionization fraction  $\bar{x}_e(z)$  as a function of redshift  $z$ .

The integral runs over all masses with a threshold mass scale  $M_{\text{th}}$ . Sheth and Tormen can be used for the halo bias

$$b_h = 1 + \frac{a\nu^2 - 1}{\delta_c} + \frac{2p}{\delta_c(1 + (a\nu^2)^p)} \quad (35)$$

Therefore, for any given model of  $\bar{x}_e(z)$  we can compute  $\bar{R}(z)$ ,  $b(z)$  and  $\bar{n}_b(z)$ . In this paper we will use  $\delta_c = 1.686$ ,  $a = 0.707$  and  $p = 0.3$  (Reed et al. 2007).

For simplicity we will assume that  $\sigma_{\ln R}$  is constant. Assuming a log normal distribution, for any given combination of  $kR(z)$  we can then read off the value of  $\langle W_R \rangle$  and  $\langle W_R^2 \rangle$  (see Figs. 12 and 13)

In Fig. 1 we have plotted the model we use for our average ionization fraction as a function of redshift. Our choice of parameters corresponds to a scenario with a neutral Universe at  $z \geq 13$  and a completely ionized Universe at  $z \leq 6$ .

## 5. TWO-BUBBLE CORRELATIONS

The cross-correlation between the neutral hydrogen and free electron fraction has two main contributions. First, the correlation is set by the Poisson distribution of ionizing sources inside the bubbles. This term is referred to as the one-bubble term, and it is dominated by the shot noise. Second, cross correlations can also be induced by the enhanced probability of bubble formation (or ionizing sources) inside overdense regions (with probability  $\langle x_e \rangle$ ). This term is referred to as the two-bubble term, and is relevant for scales much larger than the average size of a bubble.

Using Eqs. (20) and (21) we can Taylor expand  $\langle x_e \rangle$  around small overdensities, to find

$$\begin{aligned} \langle x_e \rangle &= 1 - e^{\ln(1-\bar{x}_e)(1+b\delta_W)} \\ &\simeq 1 - (1 - \bar{x}_e) [1 + b\delta_W \ln(1 - \bar{x}_e)] + \mathcal{O}(\delta_W^2) \end{aligned} \quad (36)$$

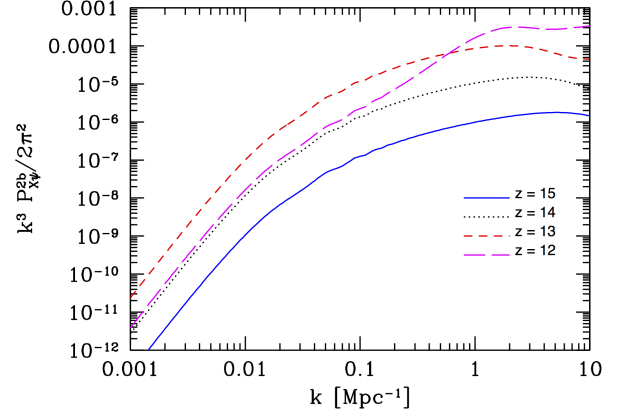
with  $b$  the bubble bias introduced earlier.

Taking the Fourier transform of  $\xi_{xx}$  and  $\xi_{x\delta}$ , we obtain

$$P_{xx}^{2b}(k) = [\bar{x}_H \ln(\bar{x}_H) b \langle W_R \rangle(k)]^2 P_{\delta\delta}(k) \quad (37)$$

$$P_{x\delta}^{2b}(k) = \bar{x}_H \ln(\bar{x}_H) b \langle W_R \rangle(k) P_{\delta\delta}(k), \quad (38)$$

Note that here we are considering correlations of the neutral hydrogen fraction (perturbing the free electron fraction accounts for a minus sign in Eq. (38)). The superscript “2b” denotes the two-bubble contribution. Also, we assume that the baryon fluctuations (the gas) trace



**Figure 2.**  $P_{X\psi}^{2b}(k)$  for 4 different redshifts with  $\sigma_{\ln R} = 0.5$ . In our toy model the two-bubble term is positively correlated for all redshifts  $z > 12$ . The zero point is uniquely determined by the relations  $\bar{x}_H = e^{-1/b}$ . Measuring this zero point could therefore help decorrelate these two parameters. However, the contribution of the two-bubble term to the overall cross correlation is very small, and it will be challenging to observe the cross correlation as a function of redshift as we show in section §7.

the dark matter fluctuations. The total two-bubble contribution to the power spectrum of  $X\psi$  results in

$$\begin{aligned} P_{X\psi}^{2b}(k) &\approx -\bar{x}_H^2 [\ln \bar{x}_H b \langle W_R \rangle(k) + 1]^2 P_{\delta\delta}(k) + \\ &\quad \bar{x}_H [\ln \bar{x}_H b \langle W_R \rangle(k) + 1] P_{\delta\delta}(k) \end{aligned} \quad (39)$$

Let us write  $P_{X\psi}^{2b}(k) = Q(1 - Q)P_{\delta\delta}(k) \equiv b_{\text{eff}}P_{\delta\delta}(k)$ , with  $Q(k, z) = \bar{x}_H \ln \bar{x}_H b \langle W_R \rangle + \bar{x}_H$ . We can distinguish two limiting cases: for  $Q > 1$  the power spectrum effective bias  $b_{\text{eff}}$  is negative, representing an *anti-correlation*, while for  $Q < 1$ ,  $b_{\text{eff}}$  is positive, and the two-bubble term is positively correlated.

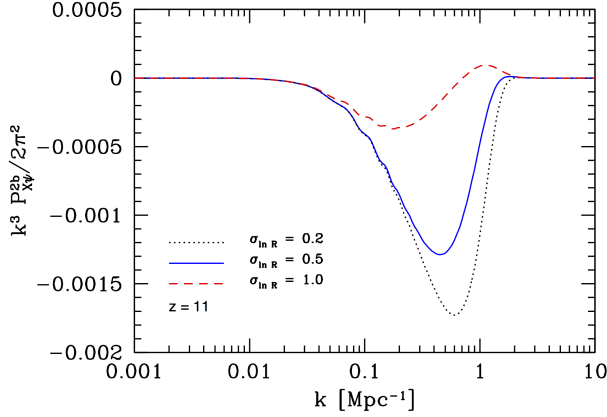
In the large scale limit, when  $k\bar{R} \ll 1$ ,  $b_{\text{eff}} \rightarrow \bar{x}_H (b \ln \bar{x}_H + 1)(\bar{x}_e - b \ln \bar{x}_H)$ . This function changes sign when  $\bar{x}_H = e^{-1/b}$ . The bubble bias on average grows towards larger  $z$ , despite the bubbles being smaller, the bubbles become rare (larger  $\bar{x}_H$ ) and highly correlated. For our choice of parameters we can solve this equation for the redshift and find  $z = 12$  as the redshift at which the two-bubble term turns negative on large scales. Fig. 2 shows the two-bubble term for various redshifts.

At scales that are smaller than the radius of the bubbles, the two-bubble term is ill-defined (Baldauf et al. 2013): the correlation length becomes shorter than the size of the bubbles, effectively rendering them to one bubble. Therefore, we will neglect the two-bubble correlation term at those scales. In practice, we apply a smoothing filter that effectively cuts the correlation for  $kR_0(z) < 3$ . We show the two-bubble contribution to the cross-correlation in Fig. 3.

## 6. ONE-BUBBLE CORRELATIONS

For scales much smaller than the average bubble radius, the correlation is dominated by the presence (or absence) of a single bubble (Wang and Hu 2006). The correlation between two (ionized) points separated by  $x_{12} = |\vec{x}_1 - \vec{x}_2|$  can be written as (Zaldarriaga et al. 2004; Furlanetto et al. 2004a)

$$\langle x_e(\vec{x}_1) x_e(\vec{x}_2) \rangle = \bar{x}_e^2 + (\bar{x}_e - \bar{x}_e^2) f(x_{12}/R), \quad (40)$$



**Figure 3.** The two-bubble term for 3 different values of the width of the log-normal distribution. The relative contribution of the two-bubble term to the total correlation increases rapidly with decreasing  $\bar{R}$  and  $\sigma_{\ln R}$ , as derived in the Appendix.

where  $f(x)$  is a function with the following limits:  $f(x) \rightarrow 1$  for  $x \ll 1$  and  $f(x) \rightarrow 0$  for  $x \gg 1$ . If the probability for finding one point inside an ionized bubble is  $\bar{x}_e$ , then when  $x_{12} \ll R$  the probability of finding the second point in the same bubble is 1, hence their joint correlation probability is  $\bar{x}_e$ . For large separations, the probability of finding two points in separate bubbles is the product of both probabilities, i.e.,  $\bar{x}_e^2$ . Eq. (40) effectively encodes the smooth transition between these two regimes. The one-bubble correlation for free electrons (or equivalently neutral hydrogen) then becomes:

$$\xi_{x_e x_e}^{1b} = \langle x_e(\vec{x}_1) x_e(\vec{x}_2) \rangle - \bar{x}_e^2 = (\bar{x}_e - \bar{x}_e^2) f(x_{12}/R) \quad (41)$$

As long as the bubbles do not overlap, the function  $f$  can be described by the convolution of two top hat window functions  $\langle W_R^2 \rangle$ . The one-bubble contribution to the power spectrum can then be written as

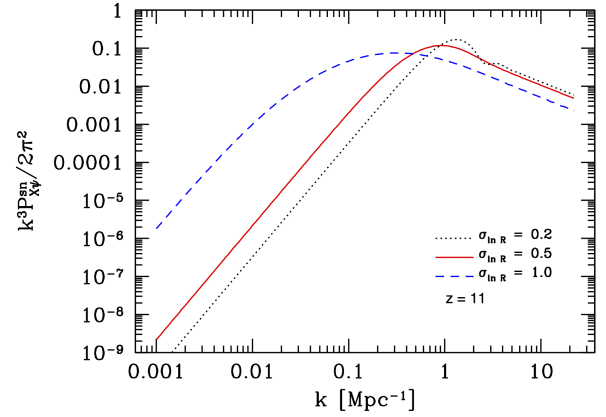
$$P_{X\psi}^{1b} = -(\bar{x}_e - \bar{x}_e^2) \left[ \langle V_b \rangle \langle W_R^2 \rangle(k) + \tilde{P}_{\delta\delta}(k) \right], \quad (42)$$

where

$$\tilde{P}_{\delta\delta}(k) = \langle V_b \rangle \int \frac{d^3 k'}{(2\pi)^3} \langle W_R^2 \rangle(k') P_{\delta\delta}(|\vec{k} - \vec{k}'|) \quad (43)$$

The first term in Eq. (42) is the shot noise of the bubbles, which is a direct consequence of randomly placing ionizing galaxies in the Universe. We will later see that this term typically dominates the total correlation function at late times. This can be understood by realizing that the bubbles tend to be larger at late times, assuming that the bubble size increases over time through bubble merging.

Note that when correlating the free electrons with the neutral hydrogen, the one-bubble contribution is always negative, i.e. these are fully anti-correlated when considering just single bubbles in the Universe. In the previous section we have also shown that the two-bubble cross-correlation at early times. As reionization proceeds and the neutral hydrogen fraction decreases, the correlation on all scales will become anti-correlated. More importantly, the signal is proportional to the matter power spectrum, which grows as  $(1+z)^2$  during matter domination. Within the model applied in this work, we find



**Figure 4.** The shot noise  $P_{X\psi}^{sn} \equiv (\bar{x}_e - \bar{x}_e^2) \langle V_b \rangle \langle W_R^2 \rangle$  for 3 different values of the width of the log-normal distribution. that at redshift  $z < 12$ , the two-bubble term is negligible compared to the one-bubble term on most scales.

It was shown by Ref. (Wang and Hu 2006) that  $\tilde{P}_{\delta\delta}(k)$  can be approximated as

$$\tilde{P}_{\delta\delta}(k) \simeq \frac{P_{\delta\delta}(k) \langle V_b \rangle \langle \sigma_R^2 \rangle}{[(P_{\delta\delta}(k))^2 + (\langle V_b \rangle \langle \sigma_R^2 \rangle)^2]^{1/2}}, \quad (44)$$

which is derived by equating the small and large scale limits of Eq. (43) with

$$\langle \sigma_R^2 \rangle = \int \frac{k^2 dk}{2\pi^2} \langle W_R^2 \rangle(k) P_{\delta\delta}(k) \quad (45)$$

We have found the simple fitting solution of Eq. (44) to be accurate to the percent level for most values of  $\{\bar{R}, \sigma_{\ln R}\}$ . In the Appendix we will show that the contribution from  $\tilde{P}_{\delta\delta}(k)$  to the one-bubble peak is relatively small for all parameter values in the range of interest for the  $\tau - 21$  cm cross-correlation, but is non-negligible and relevant at small scales.

We show the shot noise and  $\tilde{P}_{\delta\delta}$  in Figs. 4 and 5. The sum of these terms gives the one-bubble power spectrum shown in Fig. 6. Note that the one-bubble term from all three possible correlations ( $XX$ ,  $\psi\psi$  and  $X\psi$ ) is equivalent up to a sign.

In the small scale limit,  $\tilde{P}_{\delta\delta}(k) = P_{\delta\delta}(k)$ , and the one-bubble term becomes:  $P_{X\psi}^{1b}(k) = -(\bar{x}_H - \bar{x}_H^2) P_{\delta\delta}(k)$ , rendering the total cross-correlation negative at these scales (at these scales we are applying a smoothing filter to the two-bubble term, so it effectively does not contribute to the total cross-correlation). In Fig. 7 we show the total cross-correlation for  $\sigma_{\ln R} = 0.5$  at  $z = 11$  (changing the redshift of the cross-correlation will predominantly affect the average ionization fraction  $\bar{x}_e$ ).

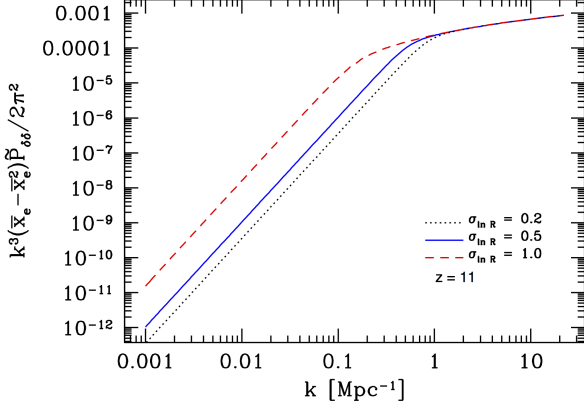
## 7. PROJECTED CROSS-CORRELATION

Fourier transforming the dimensionless brightness temperature  $\psi$ , we can write the spherical harmonic coefficient for the 21 cm fluctuation as

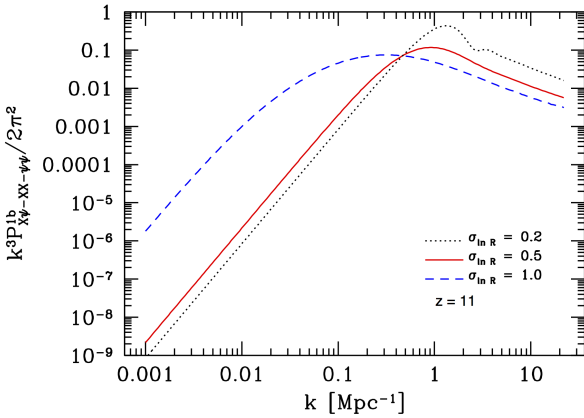
$$a_{\ell m}^{21} = 4\pi(-i)^\ell \int \frac{d^3 k}{(2\pi)^3} \hat{\psi}(\vec{k}) \alpha_\ell^{21}(k, z) Y_{\ell m}^*(\hat{k}), \quad (46)$$

where

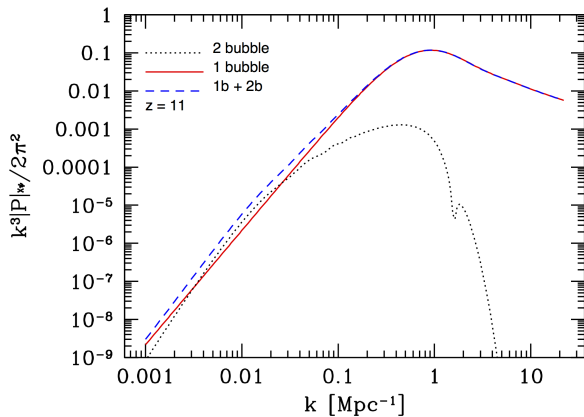
$$\alpha_\ell^{21}(k, z) = T_0(z) \int_0^\infty d\chi' W_{\chi(z)}(\chi') j_\ell(k\chi') \quad (47)$$



**Figure 5.**  $\bar{P}_{\delta\delta}$  as defined for 3 different values of the width of the log-normal distribution. This figure shows that this term is only relevant at small scales and does not contribute to the peak of the total correlation function.



**Figure 6.** The sum of the previous two figures, the total one-bubble term. The one-bubble term has the same shape for correlations  $XX$ ,  $X\psi$  and  $\psi\psi$ .



**Figure 7.**  $k^3 P_{X\psi}(k)/2\pi^2$  with  $\sigma_{\ln R} = 0.5$ .

Note that the response function is centered around  $\chi(z) = \chi'$ , and in practice we take this distance to be somewhere between  $z = 0$  and  $z = 30$ .

We can do the same for the optical depth to reioniza-

tion, i.e. :

$$\tau_{\ell m} = 4\pi(-i)^\ell \int \frac{d^3 k}{(2\pi)^3} X(\vec{k}) \alpha_\ell^\tau(k) Y_{\ell m}^*(\hat{k}), \quad (48)$$

with

$$\alpha_\ell^\tau(k) = (1 - Y_p) \sigma_T \frac{\rho_{b,0}}{m_p} \int_0^{\chi_*} \frac{d\chi'}{a^2} j_\ell(k\chi'), \quad (49)$$

where  $\chi_*$  corresponds to the distance to last scattering. Cross-correlating the two maps yields

$$\begin{aligned} \langle \tau_{\ell m} \alpha_{\ell' m'}^{21*}(z) \rangle &= \delta_{\ell\ell'} \delta_{mm'} C_\ell^{\tau, 21}(z) \\ &= \int \frac{dk}{k} \Delta_{X\psi}^2(k) \alpha_\ell^\tau(k) \alpha_{\ell'}^{21}(k, z) \end{aligned} \quad (50)$$

Here,  $\Delta_{X\psi}^2 = k^3 P_{X\psi}/(2\pi^2)$ .

Let us consider the cross-correlation in the Limber approximation. Under this approximation, we can assume that the Bessel functions are small,  $j_\ell(x) \ll 1$ , for  $x < \ell$  and peak when  $x \sim \ell$ . The integral over comoving momentum  $k$  will get most of its contribution from modes  $k \sim \ell/\chi$ . Therefore we can make the approximation that  $\Delta_{X\psi}^2(k) \sim \Delta_{X\psi}^2(\ell/\chi)$  and re-write Eq. (50) as

$$\begin{aligned} C_\ell^{\tau, 21}(z) &= (1 - Y_p) \frac{T_0(z) \rho_{b,0} \sigma_T}{m_p} \\ &\int_0^{z_*} \frac{dz'_1}{H(z'_1)} (1 + z'_1)^2 \int_0^\infty \frac{dz'_2}{H(z'_2)} W_z(\chi(z'_1)) \\ &\times 4\pi \int \frac{dk}{k} \Delta_{X\psi}^2(k) j_\ell(k\chi(z'_1)) j_\ell(k\chi(z'_2)) \end{aligned} \quad (51)$$

Again, in practice we take the window function to be centered around  $0 \leq z \leq 30$ .

In the Limber approximation we can perform the  $k$  integral over the product of Bessel functions as:

$$\int_0^\infty dk k^2 j_\ell(k\chi(z_1)) j_\ell(k\chi(z_2)) = \frac{\pi}{2} \frac{\delta(\chi(z_1) - \chi(z_2))}{\chi^2} \quad (52)$$

Thus, we can write the angular cross spectrum as

$$\begin{aligned} C_\ell^{\tau, 21}(z) &= (1 - Y_p) \frac{T_0(z) \rho_{b,0} \sigma_T}{m_p} \int_0^\infty dz' \frac{W_z(\chi(z'))}{H^2} \\ &\times \left| \frac{d\chi}{dz} \right|^{-1} \left( \frac{1 + z'}{\chi(z')} \right)^2 P_{X\psi} \left( \frac{\ell}{\chi(z')}, z' \right) \end{aligned} \quad (53)$$

We will assume that the window function is a Gaussian centered around redshift  $z$  with width  $\delta\chi$  given by

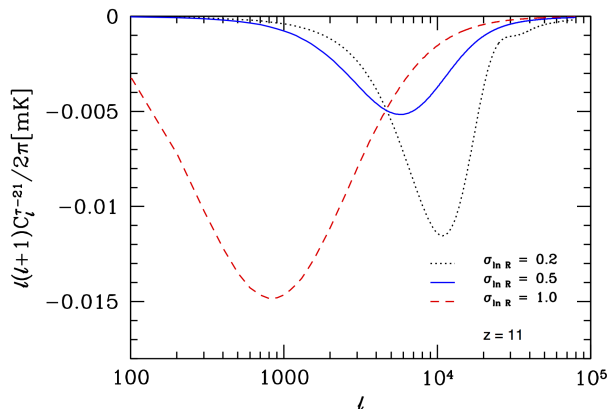
$$\delta\chi \simeq \left( \frac{\Delta\nu}{0.1 \text{ MHz}} \right) \left( \frac{1+z}{10} \right)^{1/2} \left( \frac{\Omega_m h^2}{0.15} \right)^{-1/2} \text{ Mpc} \quad (54)$$

where  $\Delta\nu$  is the bandwidth frequency of the instrument. We have taken into account that the power spectrum explicitly depends on redshift.

We show the angular cross-correlation for  $\Delta\nu = 0.2$  MHz and several values of  $\sigma_{\ln R}$  in Fig. 8.

In the previous sections we have shown that the cross-correlation between free electrons and neutral hydrogen





**Figure 8.** The cross-correlation between 21 cm temperature brightness fluctuations and the optical depth  $\tau$  at  $z = 11$ . Here we used a Gaussian window function with bandwidth frequency  $\Delta\nu = 0.2$  MHz.

has a strong dependence on the parameters that determine the bubble distribution as well as its bias. The location of the peak is set by an effective scale, which we derive in the Appendix. We will see that for a log-normal bubble distribution, this effective scale is exponential in the width of the distribution and inversely proportional to the average bubble radius. Therefore, a small change in the width of the distribution can produce a large change in the location of the peak. We have also shown that at early times the two-bubble term can be positively correlated at large scales. The positive contribution to the correlation function at large scales eventually vanishes when the universe further reionizes. However, if the bubbles are small enough, a positive contribution to the correlation function could persist until late times. Oppositely, if bubbles are relatively large (a few Mpc), the shot noise, which is negative for all scales, will dominate the correlation function. For the reionization parameterization in this paper, the shot noise is the dominated term at the most relevant redshifts (peaking around  $\bar{x}_e = 0.5$ ).

## 8. IS THE CROSS-CORRELATION DETECTABLE?

### 8.1. Signal-to-noise

In this section we will determine if the cross-correlation is detectable. An important issue that we will address here are the foregrounds. As previously mentioned, the 21 cm emission should be swamped by foregrounds, dominated on large scales by polarized Galactic synchrotron, with a total intensity of 3-4 orders of magnitude larger than the 21 cm brightness from reionization. On small scales, the redshifted 21 cm brightness is obscured by extragalactic sources (Shaver et al. 1999; Jelić et al. 2008). We do not know the spectral dependence of all these foregrounds, but in general we can assume that they are relatively smooth in frequency along the line of sight, as they are associated with same source (e.g. our own Galaxy). In principle, one can therefore remove a large part of the (large scale) foregrounds by removing the largest modes along the line of sight (see e.g. Liu and Tegmark (2012) for a recent discussion).

However, when cross-correlating the 21 cm field with the optical depth, we want to keep the largest modes, to which the integrated optical depth is most sensitive.

Hence, we will keep the foregrounds in the observed maps and show that the cross-correlation between foregrounds in the 21 cm field and in the CMB should be small. In order to neglect the cross-correlation of the foregrounds between  $\tau$  and 21 cm, we typically need the foreground of the CMB to be  $\leq 10^{-5}$  times the signal (Liu et al. 2009) (given that  $a_{\ell m}^{f,21} \sim 10^5 a_{\ell m}^{21}$ ). The synchrotron emission is roughly equal to the CMB signal at 1 GHz. Therefore, if we assume that the synchrotron scales as  $\nu^{-3}$  (Kogut et al. 2007), at 94 GHz (W band) we have  $a_{\ell m}^{\text{synchrotron}} \sim \times 10^{-6} a_{\ell m}^{\text{CMB}}$ . Thus, we estimate that the signal will be larger than the remaining foregrounds after cross-correlating the two maps.

Additionally, by not removing the foregrounds, the 21cm foregrounds will effectively act as noise term in the cross-correlation. In other words, even in the absence of correlation between foregrounds, there is still a finite probability that any given data point in the  $\tau$  map will correlate with a foreground measurement from 21 cm, i.e. the induced noise contains a term  $\langle \tau_{\ell m} a_{\ell' m'}^{f,21} \rangle$ , where the latter is the spherical harmonic coefficient of the 21 cm foreground map.

Unfortunately, we do not know exactly what the level of synchrotron foreground is, but typically  $C_{\ell}^f \sim k\ell^{-\alpha}$ , with  $2 < \alpha < 4$ . We will assume that the synchrotron emission scales as  $\nu^{-3}$ . La Porta et al. (2008) showed synchrotron emission at 480 MHz has a normalized amplitude of  $100 \text{ mK}^2 < C_{\ell=100}^f < 10000 \text{ mK}^2$ , with the actual amplitude and slope depending on the position in the sky.

Although we cannot remove the foregrounds through implementing a large scale cutoff, we can alternatively try to remove a substantial part of galactic foreground emission. If there is a large correlation between different frequencies of the foreground maps, one could measure the foregrounds at high frequency (corresponding to a completely ionized universe and, hence, with no signal in the cross-correlation), extrapolate with an appropriate scaling  $\sim \nu^{-3}$  and subtract those from the high redshift maps<sup>2</sup> (Shaver et al. 1999). If the correlation between different maps at high frequencies is of order 0.9 – 0.99, one could reduce the overall amplitude of the foreground by a factor of 10 – 100 and the power by a factor of a 100 –  $10^4$ <sup>3</sup>. In addition, Liu et al. (2012) showed that down weighting the most heavily contaminated regions in the sky can reduce the effective foreground as much as a factor of 2. Note that this approach is different from the usual spectral fitting techniques (Shaver et al. 1999; Santos et al. 2005; Wang et al. 2006; McQuinn et al. 2006; Harker et al. 2009; Jelić et al. 2010; Chapman et al. 2012, 2013).

At small scales we expect extra galactic radio sources to dominate the foregrounds. However, there are several strategies that will likely suppress the noise term

<sup>2</sup> This scaling is approximate and simplistic. In reality one would probably have to consider a slope that changes as a function of scale and frequency. We are assuming the scaling will be further understood as a function of frequency once we are capable of performing this cross correlation.

<sup>3</sup> Note that this is a very crude estimate. For example, it might be relevant to consider 21 cm signals after reionization (at low  $z$ ) due to residual neutral hydrogen, primarily in Damped Ly $\alpha$  absorbers (DLA's).

due to correlations between millimeter and radio emission: (i) since bright sources are expected to dominate the variance at radio frequency (de Zotti et al. 2010; Planck Collaboration et al. 2011; Heywood et al. 2013), these sources can be masked at  $5\sigma$ . This will suppress the radio source contribution without removing very much of the sky; (ii) at a given location in the map, the radio data short wards and long-wards of the 21 cm radio emission can be used to remove both galactic and extragalactic foreground by assuming that the sources at a given location can be fit by a power law; (iii) at millimeter and sub millimeter wavelengths, multi-frequency data can be used to separate CMB signal from dusty galaxy foregrounds. The Planck data shows that the 353 GHz data can be used to remove  $> 90\%$  of the dusty galaxy foreground at 220 GHz (Planck Collaboration et al. 2013). All of these strategies will likely be employed to remove foregrounds in both maps.

We will consider a case in which the angular power spectrum of the foreground is given by:

$$C_\ell^f(z) \simeq \frac{10^6}{2} \frac{1}{c_f} \text{mK}^2 \ell^{-3} \left( \frac{f(z)}{480 \text{ MHz}} \right)^{-3}, \quad (55)$$

where  $f(z)$  corresponds to the frequency of the redshift considered and  $100 \leq c_f \leq 10^4$  is the foreground reduction factor that we can hope to achieve through a measurement at low redshift. As we will show later, the signal-to-noise of the cross-correlation does not vary substantially for different values of  $c_f$  in this range.

We will assume a noise power spectrum given by (Morales 2005; McQuinn et al. 2006; Adshead and Furlanetto 2007; Mao et al. 2008),

$$N_\ell^{21,21} = \frac{2\pi}{\ell^2} (20 \text{ mK})^2 \left[ \frac{10^4 \text{ m}^2}{A_{\text{eff}}} \right]^2 \left[ \frac{10'}{\Delta\Theta} \right]^4 \left[ \frac{1+z}{10} \right]^{9.2} \times \left[ \frac{\text{MHz}}{\Delta\nu} \frac{100 \text{ hr}}{t_{\text{int}}} \right] \quad (56)$$

We will do forecasts for a total integration time of 1000 hours, and a beam with an angular diameter of  $\Delta\Theta = 9$  arcmin. We set the bandwidth to  $\Delta\nu = 0.2$  MHz. For a LOFAR (van Haarlem et al. 2013) type experiment, we use  $A_{\text{eff}} = 10^4 \text{ m}^2$  and for a Square Kilometer Array (SKA) type experiment we use  $A_{\text{eff}} = 10^5 \text{ m}^2$ .

On the CMB side, experiments are rapidly improving (Pla 2006; Niemack et al. 2010; Austermann et al. 2012; Zaldarriaga et al. 2008), with high sensitivity experiments coming soon (Planck, ACTPol, SPTPol, CMBPol) and we should have observations of the E- and B-modes polarization spectra up to small scales in the near future. We will now consider a next generation polarization experiment that allows us to reconstruct a map of the optical depth  $\tau_{\ell m}$  with the estimator proposed by Dvorkin and Smith (2009). This estimator was built to extract the inhomogeneous reionization signal from future high-sensitivity measurements of the cosmic microwave background temperature and polarization fields. Dvorkin and Smith (2009) wrote a minimum variance quadratic estimator for the modes of the optical depth field given by:

$$\hat{\tau}_{\ell m} = N_\ell^{\tau\tau} \sum_{\ell_1 m_1 \ell_2 m_2} \Gamma_{\ell_1 \ell_2 \ell}^{EB} \begin{pmatrix} \ell_1 & \ell_2 & \ell \\ m_1 & m_2 & m \end{pmatrix} \times \frac{a_{\ell_1 m_1}^{E*} a_{\ell_2 m_2}^{B*}}{(C_{\ell_1}^{EE} + N_{\ell_1}^{EE})(C_{\ell_2}^{BB} + N_{\ell_2}^{BB})}, \quad (57)$$

where  $C_\ell^{EE}$  and  $C_\ell^{BB}$  are the E- and B-mode polarization power spectra.  $N_\ell^{EE}$  and  $N_\ell^{BB}$  correspond to the CMB noise power spectra, and are given by:

$$N_\ell^{EE} = N_\ell^{BB} = \Delta_P^2 \exp\left(\frac{\ell(\ell+1)\theta_{\text{FWHM}}^2}{8 \ln(2)}\right), \quad (58)$$

where  $\Delta_P$  is the detector noise and  $\theta_{\text{FWHM}}$  is the beam size.

The coupling  $\Gamma_{\ell_1 \ell_2 \ell}^{EB}$  can be written as

$$\Gamma_{\ell_1 \ell_2 \ell}^{EB} = \frac{C_{\ell_1}^{E_0 E_1}}{2i} \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell+1)}{4\pi}} \times \left[ \begin{pmatrix} \ell_1 & \ell_2 & \ell \\ -2 & 2 & 0 \end{pmatrix} - \begin{pmatrix} \ell_1 & \ell_2 & \ell \\ 2 & -2 & 0 \end{pmatrix} \right] \quad (59)$$

Here  $C_{\ell_1}^{E_0 E_1}$  is the cross-power spectrum between the CMB  $E_0$ -mode polarization without patchy reionization and the response field to  $\tau$  fluctuations  $E_1$ . ( $C_{\ell_1}^{E_0 E_1}$  is positive at large scales due to Thomson scattering, and negative at small scales due to the screening).

Furthermore, the reconstruction noise power spectrum is given by:

$$N_\ell^{\tau\tau} = \left[ \frac{1}{2\ell+1} \sum_{\ell_1 \ell_2} \frac{|\Gamma_{\ell_1 \ell_2 \ell}^{EB}|^2}{(C_{\ell_1}^{EE} + N_{\ell_1}^{EE})(C_{\ell_2}^{BB} + N_{\ell_2}^{BB})} \right]^{-1} \quad (60)$$

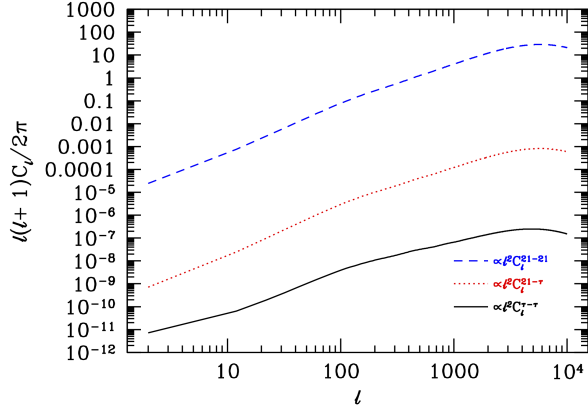
We note that the main source of contamination in reconstructing  $\tau(\hat{n})$  comes from the non-Gaussian signal from gravitational lensing of the CMB. In principle, unbiased estimators that simultaneously reconstruct the inhomogeneous reionization signal and the gravitational potential can be constructed (Su et al. 2011). For purposes of simplicity, we will estimate our cross-correlation using the estimator given by Eq. (57), but the results in this work are straightforward to generalize.

Given that the  $\tau$  map is not sensitive to redshift (Dvorkin and Smith (2009) showed that the estimator is only sensitive to one principal component in redshift), while the 21 cm map can be reconstructed on redshift slices, we will give a weight to the cross-correlation. This weight will be built in order to maximize the signal to noise, in the same spirit as in Peiris and Spergel (2000), where a weight was derived for the cross-correlations between CMB and Galaxy surveys. We write the weighted 21 cm maps as

$$\tilde{a}_{\ell m}^{21}(z) = \int_0^z dz' a_{\ell m}^{21}(z') w_\ell(z') \quad (61)$$

We then want to maximize

$$\chi_{\ell m}^2 = \frac{\langle \tau_{\ell m}^* \tilde{a}_{\ell m}^{21}(z) \rangle^2}{\langle \tau_{\ell m} \tau_{\ell m}^* \rangle \langle \tilde{a}_{\ell m}^{21}(z) \tilde{a}_{\ell m}^{21*}(z) \rangle}, \quad (62)$$



**Figure 9.** The different power spectra used to compute the signal to noise in Fig. 10. The sign of  $C_\ell^{21-\tau}$  has been inverted for the sake of comparison. The spectra are shown at  $z = 11$ . The bubble radii as a function of redshift, the bubble bias and the bubble number density are all determined through our toy reionization model of Eq. (22) as explained in section §4. The noise to the  $\tau$  estimator is given by Eq. (60) (Dvorkin and Smith 2009).

and, in doing so, we find:

$$w_\ell(z) = \frac{C_\ell^{\tau,21}(z)}{(C_\ell^{21,21} + N_\ell^{21,21} + C_\ell^f)(z)}, \quad (63)$$

which is nothing else then the projected signal over the projected noise. Note that we have included the foreground as a source of noise as explained before.

We can now compute the signal to noise for the  $\tau$ -21 cm cross-correlation as

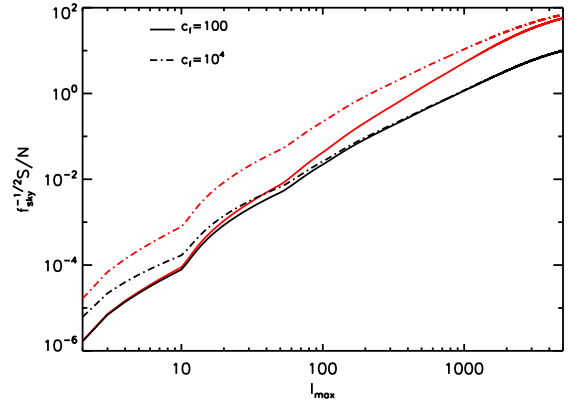
$$\left(\frac{S}{N}\right)^2 = f_{\text{sky}} \sum_\ell (2\ell + 1) \times \int dz \frac{|C_\ell^{\tau,21}(z)|^2}{(C_\ell^{\tau,\tau} + N_\ell^{\tau\tau})(C_\ell^{21,21} + N_\ell^{21,21} + C_\ell^f)(z)} \quad (64)$$

In Fig. 10 we assess the level of detectability for a reionization history with  $\sigma_{\ln R} = 0.5$  (see Fig. 9 for an example angular spectra at  $z = 11$ ). Again, to generate the spectra we use a Gaussian window function with  $\Delta\nu = 0.2$  MHz. We find that the signal to noise reaches  $f_{\text{sky}}^{-1/2} S/N = 0.2$  at  $\ell_{\text{max}} = 3000$  for a LOFAR type experiment (with  $f_{\text{sky}} = 0.001$ ), while for a SKA type experiment (with  $f_{\text{sky}} = 0.25$ ) it reaches  $f_{\text{sky}}^{-1/2} S/N = 16$ , with  $f_{\text{sky}}$  being the fraction of the sky covered. We note that at small scales, the signal to noise does not vary substantially when considering different values of the parameter  $c_f$ , that represents the level of foreground subtraction.

## 8.2. Reionization parameters

In this section we will assess what we can learn about reionization by studying the 21 cm- $\tau$  cross-correlation. We will forecast parameter uncertainties in the following parameters  $\pi = \{\tau, \Delta y\}$ .

As a forecasting tool we will use a Fisher matrix analysis, where the Fisher matrix is given by (Tegmark et al.



**Figure 10.** Total signal-to-noise for the 21 cm- $\tau$  cross-correlation as a function of  $\ell_{\text{max}}$  for a model with a log normal bubble distribution with a width  $\sigma_{\ln R} = 0.5$ . We consider an experiment with CMB noise power spectra given by  $\Delta_P = 0.3 \mu\text{K-arcmin}$  and beam size  $\Theta_{FWHM} = 1$  arcmin. The foreground angular power spectrum is given by Eq. (55). We show forecasts for an experiment with the planned noise level of SKA (in red lines) and the planned noise level of LOFAR (in black lines). Note that the y-axis represents the product of the signal-to-noise and the fraction of the sky covered. For an experiment like LOFAR (with  $f_{\text{sky}} = 0.001$ ), we find  $S/N = 0.2$  and for SKA (with  $f_{\text{sky}} = 0.25$ ), we get  $S/N = 16$  at  $\ell_{\text{max}} = 3000$  with  $c_f = 100$ . At small scales does not vary substantially when considering different values of the parameter  $c_f$ , which represents the level of foreground subtraction.

1997):

$$F_{\mu\nu} = f_{\text{sky}} \sum_\ell (2\ell + 1) \times \int dz \frac{(\partial C_\ell^{\tau,21}(z)/\partial \pi_\mu)(\partial C_\ell^{\tau,21}(z)/\partial \pi_\nu)}{(C_\ell^{\tau,\tau} + N_\ell^{\tau\tau})(C_\ell^{21,21} + N_\ell^{21,21} + C_\ell^f)(z)} \quad (65)$$

where  $\mu$  and  $\nu$  run over the parameter modes.

The rms uncertainty on the parameter  $\pi_\mu$  is given by  $\sigma(\pi_\mu) = (F_{\mu\mu}^{-1})^{1/2}$  if the other parameters are marginalized. If the remaining parameters are assumed fixed, then the rms is  $\sigma(\pi_\mu) = (F_{\mu\mu})^{-1/2}$ .

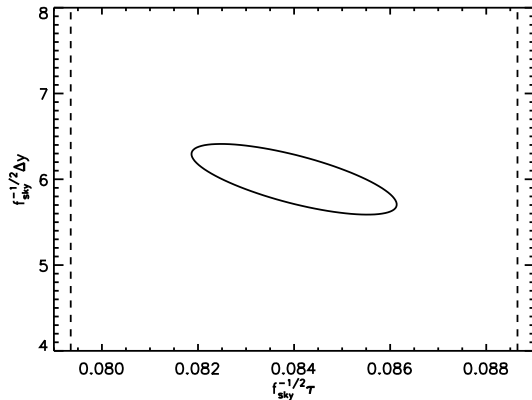
We consider a next generation polarization experiment with noise power spectrum given by  $\Delta_P = 0.3 \mu\text{K-arcmin}$  and beam size  $\Theta_{FWHM} = 1$  arcmin.

When assuming an experiment with the planned noise level of SKA and  $f_{\text{sky}} = 0.25$ , the width of reionization  $\Delta y$  can be constrained at the 10% level, and  $\tau$  at the 4% level, when the remaining parameters are considered fixed.

We show the error ellipses for the optical depth and the width of reionization in Fig. 11. The Planck priors are shown in dashed lines.

## 9. DISCUSSION AND CONCLUSION

We investigated the correlation between free electrons, traced by the optical depth  $\tau$ , and neutral hydrogen, traced through the emission of 21 cm photons, during the epoch of reionization. To compute the cross-correlation we used a simple model where patches of ionized gas are represented by spherical bubbles. The cross-correlation will depend on the presence or absence of these bubbles (the one-bubble term) and the clustering of bubbles (the



**Figure 11.** Forecasted uncertainties on the width of reionization parameter  $\Delta y$  and the optical depth  $\tau$  (assuming that the other parameters are fixed) for an experiment with the planned noise level of SKA. Note that the y-axis represents the product of the parameter value and the fraction of the sky observed. For reference, the dashed lines correspond to the Planck priors on the optical depth.

two-bubble term). As expected, the cross-correlation is negative on small scales, where it is dominated by the shot noise of the bubbles. On large scales, the two-bubble term can render the correlation positive as long as the effective bias  $b_{\text{eff}}$  is large or  $\bar{x}_e$  is small. Small bubbles at a fixed neutral hydrogen fraction imply a small bubble bias, hence the two-bubble term has a suppressed (positive) amplitude. A larger correlation could be driven by the ionization fraction, but within a bubble merger scenario the smallest bubbles are expected at early times, when the ionization fraction is small and the total matter power spectrum is suppressed.

The anti-correlation peak, set by the sum of the one and the two-bubble terms, depends critically on the distribution of the ionized bubbles. Consequently, a measurement of the cross-correlation allows us to probe the parameters relevant for the reionization history. In principle a measurement of a positive correlation at early times, would theoretically allow us to entangle the degeneracy between  $\tau$  and the bubble bias. However, we showed that the two-bubble term typically has a very small amplitude.

One major obstacle in measuring the 21 cm emission from the EoR are the large foregrounds at these frequencies. For the auto-correlation, any detection requires a careful removal of foregrounds, which typically results in the removal of the largest modes along the line of sight. The advantage of the cross-correlation is that foregrounds in the measurement of  $\tau$  are weakly correlated with those in the 21 cm field. Therefore, the cross-correlation is less sensitive to the detailed understanding of the foregrounds.

In this paper we have computed the signal to noise of the cross-correlation using the estimator for inhomogeneous reionization  $\hat{\tau}_{\ell m}$  proposed by Dvorkin and Smith (2009). In our computation there is very little contribution from any positive correlation at large scales coming from two-bubble term, and most of the signal comes from the shot noise. Because a measurement of the optical depth gets most of its signal from the long wavelength mode along the line of sight, we left the 21 cm foregrounds as a noise term. Although the signal to noise per mode is small, the large number of modes allows for a detection when considering a next generation 21 cm experiment cross-correlated with a CMB experiment that measures the polarization  $B$ -modes in most of the sky. We expect that around the time SKA observes a large part of the sky, CMB experiments will have improved to the level that we are able to reconstruct a map of  $\tau_{\ell m}$ . Although the auto-correlation of both maps will give significant insight into reionization, cross-correlating these maps will provide us with a complementary probe. We find that a measurement of this cross-correlation with a detector noise level of SKA (and  $f_{\text{sky}} = 0.25$ ) on the 21cm side and noise level of a next generation polarization type experiment on the CMB side constrains the width of the ionization history at the 10% level and the optical depth at the 4% level.

The authors would like to thank Renyue Cen, Enrico Pajer, Fabian Schmidt, Kendrick Smith, and Matias Zaldarriaga for useful discussions. P.D.M. is supported by the Netherlands Organization for Scientific Research (NWO), through a Rubicon fellowship. C.D. is supported by the National Science Foundation grant number AST-0807444, NSF grant number PHY-0855425, and the Raymond and Beverly Sackler Funds. P.D.M. and D.N.S. are in part funded by the John Templeton Foundation grant number 37426.

## APPENDIX

### REIONIZATION MODEL DEPENDENCE

The choice of a log-normal distribution is motivated in part by simulations in (Zahn et al. 2006) and (Wang and Hu 2006). In this appendix we derive constraints on the relative contributions of the various terms of the cross correlation as a function  $b$ ,  $\bar{R}$  and  $\sigma_{\ln R}$ . The aim of this appendix is to show that in most realistic scenarios, the shot noise is generally the dominating term, independent of reionization details.

#### Log-normal distribution

We first start by investigating the implications of a log-normal distribution for the bubble radius. The bubble distribution is given by

$$P(R, \sigma_{\ln R}) = \frac{1}{R} \frac{1}{\sqrt{2\pi\sigma_{\ln R}^2}} e^{-[\ln(R/\bar{R})]^2 / (2\sigma_{\ln R}^2)} \quad (\text{A1})$$



Given this distribution we can compute the average bubble size:

$$\langle V_b \rangle = \int dR P(R) V_b(R) = \frac{4\pi \bar{R}^3}{3} e^{9\sigma_{\ln R}^2/2} \quad (\text{A2})$$

To address the dependence of the resulting correlation function, we also need the variance

$$\langle V_b^2 \rangle = \int dR P(R) V_b^2(R) = \frac{(4\pi)^2 \bar{R}^6}{9} e^{18\sigma_{\ln R}^2} \quad (\text{A3})$$

The amplitude of the one and two-bubble terms, and the relevant scale where these peak, strongly depend on the window function

$$W_R(k) = \frac{3}{(kR)^3} [\sin(kR) - kR \cos(kR)] \quad (\text{A4})$$

Recall the the volume averaged window function and window function squared (shown in Figs. 12 and 13) are defined as

$$\langle W_R \rangle(k) = \frac{1}{\langle V_b \rangle} \int_0^\infty dR P(R) V_b(R) W_R(kR), \quad (\text{A5})$$

and

$$\langle W_R^2 \rangle(k) = \frac{1}{\langle V_b \rangle^2} \int_0^\infty dR P(R) V_b^2(R) W_R^2(kR) \quad (\text{A6})$$

The total correlation function can be written as

$$P_{X\psi} = P_{X\psi}^{1b} + P_{X\psi}^{2b}, \quad (\text{A7})$$

where the one-bubble contribution consists of two relevant terms: the shot noise and the power spectrum  $\tilde{P}_{\delta\delta}$  given by Eq. (44).

Let us define the following relevant ratios:

$$R^{2b-sn} \equiv P_{X\psi}^{2b}/P_{X\psi}^{sn} \quad (\text{A8})$$

and

$$R^{1b-sn} \equiv \tilde{P}_{\delta\delta}/P_{X\psi}^{sn}, \quad (\text{A9})$$

where  $P_{X\psi}^{sn}$  is the contribution from the shot noise, which is given by

$$P_{X\psi}^{sn} = -(\bar{x}_e - \bar{x}_e^2) \langle V_b \rangle \langle W_R^2 \rangle(k) \quad (\text{A10})$$

In the limit of small comoving momenta (large scales),  $\langle W_R \rangle \rightarrow 1$ , and  $\langle W_R^2 \rangle \rightarrow \langle V_b^2 \rangle / \langle V_b \rangle^2$ . For small scales, we have:

$$\begin{aligned} \langle W_R^2 \rangle(k, \bar{R}, \sigma_{\ln R}) &\sim \frac{9}{2k^4 \langle V_b \rangle^2} \int_0^\infty \frac{dR}{R^4} P(R) V_b^2(R) \\ &= \frac{9}{2k^4 \bar{R}^4} e^{-7\sigma_{\ln R}^2} \end{aligned} \quad (\text{A11})$$

Roughly speaking, we know that the contribution from the shot noise term will be constant and have a peak at some characteristic scale after which it will decrease as  $\propto 1/k^4$ . Furthermore, up until that characteristic scale, the amplitude of the shot noise is boosted with respect to all the other terms as  $\langle V_b^2 \rangle / \langle V_b \rangle^2 = e^{9\sigma_{\ln R}^2}$ .

The characteristic scale is determined by equating the two limiting cases (Mortonson and Hu 2007), i.e.

$$\frac{(4\pi)^2 \bar{R}^6}{9} e^{18\sigma_{\ln R}^2} = \frac{(4\pi)^2 \bar{R}^2}{2k^4} e^{2\sigma_{\ln R}^2}, \quad (\text{A12})$$

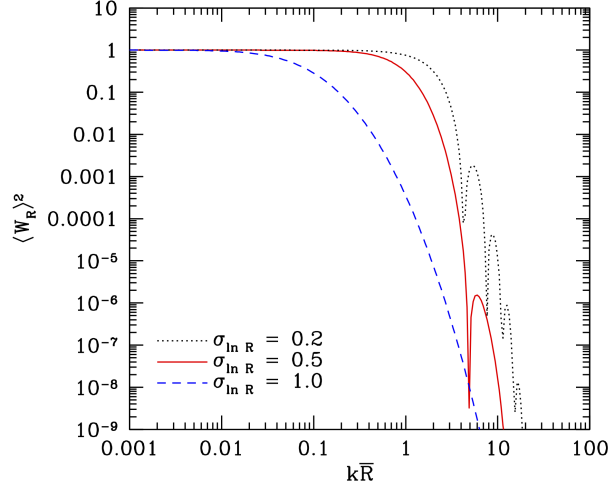
This tells us that the shot noise roughly peaks around

$$k_{peak} = \left( \frac{9}{2\bar{R}^4} e^{-16\sigma_{\ln R}^2} \right)^{1/4} \quad (\text{A13})$$

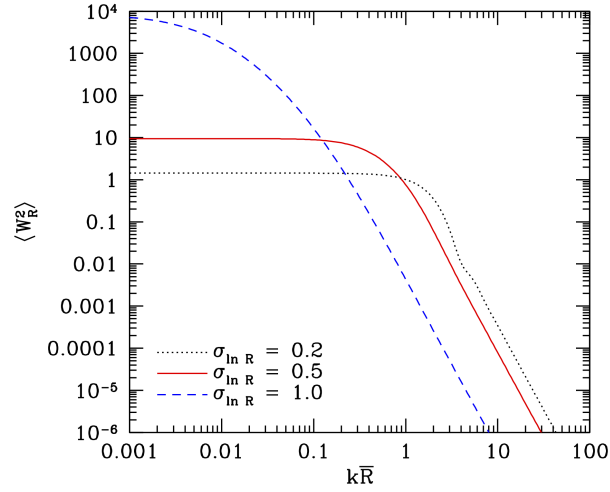
The second contribution to the one-bubble term comes from  $\tilde{P}_{\delta\delta}(k)$ , which is given by

$$\tilde{P}_{\delta\delta}(k) = \langle V_b \rangle \int \frac{d^3 k'}{(2\pi)^3} \langle W_R^2 \rangle(k') P_{\delta\delta}(|\vec{k} - \vec{k}'|) \quad (\text{A14})$$





**Figure 12.**  $\langle W_R \rangle^2$  for different values of  $\sigma_{\ln R}$ .



**Figure 13.**  $\langle W_R^2 \rangle$  for different values of  $\sigma_{\ln R}$ . Note that the maximum amplitude in the limit  $k\bar{R} \ll 1$  grows exponentially in  $\sigma_{\ln R}$ , which will be relevant for the relative contribution of the shot noise with respect to the one and two-bubble terms.

As it was shown by Wang and Hu (2006), in the large scale limit we have:

$$\lim_{k\bar{R} \ll 1} \tilde{P}_{\delta\delta}(k) \simeq \langle V_b \rangle \int_0^\infty \frac{dk}{2\pi^2} k^2 \langle W_R^2 \rangle(k) P_{\delta\delta}(k) \quad (\text{A15})$$

In this limit,  $\langle W_R^2 \rangle \rightarrow \langle V_b^2 \rangle / \langle V_b \rangle^2$ , which allows us to put the following constraint on the amplitude of  $\tilde{P}_{\delta\delta}(k)$

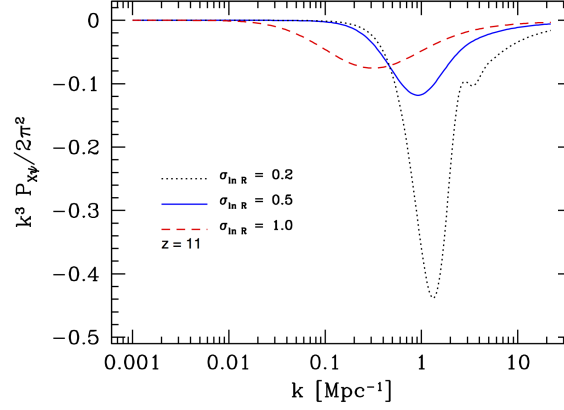
$$\tilde{P}_{\delta\delta} \lesssim \frac{\langle V_b^2 \rangle}{\langle V_b \rangle} \int_0^{k_{peak}} k^2 dk P_{\delta\delta}(k) \quad (\text{A16})$$

The relative peak amplitude between the two contributions in the one-bubble term is therefore given by

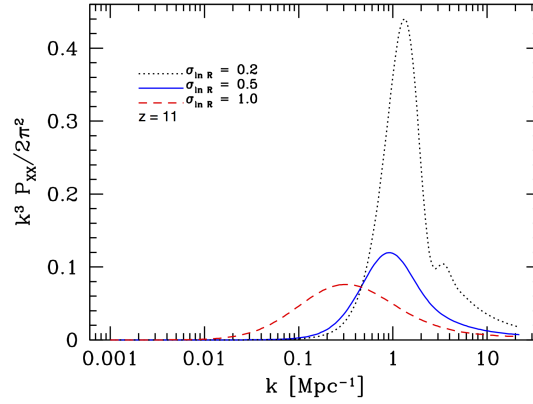
$$R^{1b-sn} \equiv \tilde{P}_{\delta\delta} / P_{X\psi}^{sn} \lesssim \frac{1}{2\pi^2} \int_0^{k_{peak}} k^2 dk P_{\delta\delta}(k) \quad (\text{A17})$$

Since  $k_{peak}$  depends on the bubble radius and on  $\sigma_{\ln R}$ , so does the relative contribution. Generally speaking, a narrower distribution (with smaller  $R_0$ ) leads to a larger contribution from  $\tilde{P}_{\delta\delta}$  to the total one-bubble term. That being said, even for very narrow distributions and very small average bubble radius we find that the total contribution to the peak does not exceed more than a few percent, i.e., in realistic scenarios the shot noise term dominates the total one-bubble term (see e.g. Zahn et al. (2010)). In Fig. 14 we plot the cross-correlation  $X\psi$  for different values of  $\sigma_{\ln R}$ , confirming our estimate for the peak sale in Eq. (A13).

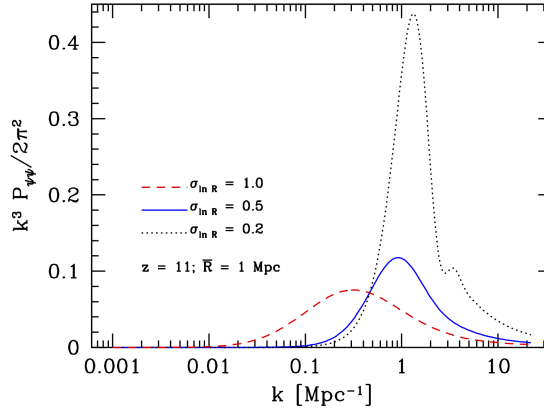
There is one caveat, which is that  $\tilde{P}_{\delta\delta}$  does not drop as fast as the shot noise, hence at small scales this term can contribute more. In fact, it is this term the responsible for the turnover at small scales of the  $\tau$ - $\tau$  and 21-21 auto



**Figure 14.**  $k^3 P_{X\psi}(k)/2\pi^2$  for  $\bar{R} = 1$  Mpc,  $b = 6$  at different values of  $\sigma_{\ln R}$ .



**Figure 15.**  $k^3 P_{XX}(k)/2\pi^2$  for  $\bar{R} = 1$  Mpc,  $b = 6$  at different values of  $\sigma_{\ln R}$ .



**Figure 16.**  $k^3 P_{\psi\psi}(k)/2\pi^2$  for  $\bar{R} = 1$  Mpc,  $b = 6$  at different values of  $\sigma_{\ln R}$ .

power spectra (see Figs. 15 and 16).

We remind the reader that the two-bubble contribution to the  $X\psi$  cross-correlation is given by:

$$P_{X\psi}^{2b} \approx -\bar{x}_H^2 [\ln \bar{x}_H b \langle W_R \rangle + 1]^2 P_{\delta\delta}(k) + \bar{x}_H [\ln \bar{x}_H b \langle W_R \rangle + 1] P_{\delta\delta}(k) \quad (\text{A18})$$

The minimum value  $P_{X\psi}^{2b}$  is reached when  $\bar{x}_H = e^{-1-1/b}$ , while for the shot noise term  $\bar{x}_e = 0.5$  represents the peak

value. We can now compare the two-bubble terms to the shot noise:

$$R^{2b-sn} \equiv P_{X\psi}^{2b}/P_{X\psi}^{sn} \lesssim \frac{16}{3\pi} b e^{-1-1/b} (1 + e^{-1-1/b}) \times P_{\delta\delta}(k_{peak}) e^{-27\sigma_{\ln R}/2} \quad (\text{A19})$$

For the toy reionization model, we find that  $\bar{R}(z=11) = 1.14$  Mpc with a  $\sigma_{\ln R} = 0.5$  and  $b(z=11) = 4.8$ , we find  $R^{2b-sn} \sim 0.01$ , close to the ratio found in Fig. 7.

Can the two-bubble term ever dominate over the one-bubble term? Since the ratio goes as  $1/\bar{R}^3$  and decreases exponentially in  $\sigma_{\ln R}$ , for bubble distributions with small bubble radius and narrow width, we find that the two-bubble term can easily dominate the total correlation function. When the radius and the variance depend on redshift, we expect the two-bubble term to be increasingly important to the correlation function in the early stages of reionization, in other words, when the correlation function is dominated by points that live in two different bubbles. At the onset of reionization, the bubbles are small and it is more probable to find two points that live in two separate bubbles. As bubbles merge and grow, it becomes more likely that the correlation function has contribution from points that are in the same bubble. McQuinn et al. (2006) make this distinction, and divide the reionization model in two regimes separated by the average ionization fraction.

Concluding, we see that the location of the peak of the correlation function is roughly set by  $k_{peak}$ , given in Eq. (A13), and we note that the location of the peak is almost equivalent for the one and two-bubble terms.

We have shown that in the case of a log-normal distribution the contribution from  $\tilde{P}_{\delta\delta}$  to the peak amplitude generally is small compared to the shot noise and as such, to the overall correlation. Since the shot noise grows as  $\propto \bar{R}^3$  and exponentially in the width of the distribution  $\sigma_{\ln R}$ , we find that assuming smaller values for these parameters lead to rapidly increasing contribution of the two-bubble term compared to the 1-bubble term. Since decreasing both of these parameters also increases the value of  $k_{peak}$ , assuming a narrower distribution of bubbles with a smaller average radius results in a correlation function that is dominated by the two-bubble term and peaks at smaller (physical) scales. These findings are consistent with the expectation that larger bubble imply a larger shot noise.

#### Normal distribution

Simulations show that bubbles are well traced by a log-normal distribution at early times (Zahn et al. 2006), while at later times the distribution can transition to a normal distribution. Since at late times, large radii dominate reionization, we expect the shot noise to become more dominant.

A normal distribution is given by

$$P(R) = \frac{1}{\sqrt{2\pi\sigma_R^2}} e^{[-(R-\bar{R})^2/(2\sigma_R^2)]} \quad (\text{A20})$$

As expected, for a Gaussian, all relevant quantities are much closer to the distribution values (e.g.  $\bar{R}$ ) deviating, by definition, at most 1 sigma.

The average bubble volume is given by

$$\langle V_b \rangle = \frac{4}{3}\pi\bar{R}(\bar{R}^2 + 3\sigma_R^2) \quad (\text{A21})$$

We will assume that a normal distribution is only valid for  $\bar{R} > 1\text{Mpc}$ . In this limit, the above equality holds, even for a bound probability function, as long as  $\sigma_R \leq 1$ . This cutoff is consistent with observations. The expression above can easily be understood by Wick expanding the 3-point function  $\langle R^3 \rangle$ .

Similarly, for the variance we obtain:

$$\langle V_b^2 \rangle = \frac{16}{9}\pi^2 (15\bar{R}^4\sigma_R^2 + 45\bar{R}^2\sigma_R^4 + \bar{R}^6 + 15\sigma_R^6) \quad (\text{A22})$$

A gaussian distribution allows us to analytically compute the volume average window function:

$$\langle W_R \rangle(k) = \frac{3e^{-\frac{1}{2}k^2\sigma_R^2} [(k^2\sigma_R^2 + 1) \sin(k\bar{R}) - k\bar{R} \cos(k\bar{R})]}{k^3 (3\bar{R}\sigma_R^2 + \bar{R}^3)}, \quad (\text{A23})$$

At large scales,

$$\lim_{k\bar{R} \ll 1} \langle W_R \rangle = 1, \quad (\text{A24})$$

while at small scales,

$$\lim_{k\bar{R} \gg 1} \langle W_R \rangle = \frac{3\sigma_R^2 e^{-\frac{1}{2}k^2\sigma_R^2} \sin(k\bar{R})}{k\bar{R} (3\sigma_R^2 + 1)} \quad (\text{A25})$$

The variance is given by:

$$\langle W_R^2 \rangle(k) = \frac{9e^{-2k^2\sigma_R^2} \left\{ e^{2k^2\sigma_R^2} [k^2 (\bar{R}^2 + \sigma_R^2) + 1] - 2k\bar{R} (2k^2\sigma_R^2 + 1) \sin(2k\bar{R}) + [k^2 (\bar{R}^2 - 3\sigma_R^2) - 4k^4\sigma_R^4 - 1] \cos(2k\bar{R}) \right\}}{2k^6 (3\bar{R}\sigma_R^2 + \bar{R}^3)^2} \quad (\text{A26})$$

At large scales, the variance becomes:

$$\lim_{k\bar{R} \ll 1} \langle W_R^2 \rangle = \langle V_b^2 \rangle / \langle V_b \rangle^2, \quad (\text{A27})$$

and at small scales:

$$\lim_{k\bar{R} \gg 1} \langle W_R^2 \rangle = 9(\bar{R}^2 + \sigma_R^2) / (2k^4(\bar{R}^2 + 3\sigma_R^2)^2) \quad (\text{A28})$$

By equating the two limiting cases, we can derive the peak scale for the variance of  $W_R$ :

$$k_{peak}^{var} = \left[ \frac{9(\bar{R}^2 + \sigma_R^2)(3\bar{R}\sigma_R^2 + \bar{R}^3)}{2(15\bar{R}^4\sigma_R^2 + 45\bar{R}^2\sigma_R^4 + \bar{R}^6 + 15\sigma_R^6)} \right]^{1/4} \quad (\text{A29})$$

The maximum peak value of the shot noise is then given by

$$k^3 P_{X\psi}^{sn} \lesssim (k_{peak}^{var})^3 \frac{\langle V_b^2 \rangle}{\langle V_b \rangle} \quad (\text{A30})$$

A similar approach for the peak scale of the window average does not give an accurate enough answer. Therefore, we will use the following best fit:

$$k_{peak}^{av} \sim 2.2\bar{R}^{-1} \quad (\text{A31})$$

This and the derived maximum value of the shot noise immediately allow us to put a constraint on the ratio between the two-bubble term and the shot noise,

$$R^{2b-sn} \equiv P_{X\psi}^{2b} / P_{X\psi}^{sn} \simeq b \frac{1 + \bar{x}_H}{1 - \bar{x}_H} \frac{P_{\delta\delta}(k_{peak}^{av})}{\langle V_b \rangle \langle W_R^2 \rangle (k_{peak}^{var})} \quad (\text{A32})$$

From this expression, we find that the shot noise is significantly larger than the two-bubble term around the peak scale. This nicely fits into the previous picture, since the normal distribution of the bubbles is only physical at late time in the reionization history when  $\bar{R} > 1$  Mpc. At those times the one-bubble term (dominated by the shot noise) should make up the largest contribution to the total cross-correlation spectrum.

## REFERENCES

- A. Loeb and R. Barkana, *Ann.Rev.Astron.Astrophys.* **39**, 19 (2001), [arXiv:astro-ph/0010467 \[astro-ph\]](#).
- X.-H. Fan, M. A. Strauss, R. H. Becker, R. L. White, J. E. Gunn, *et al.*, *Astron.J.* **132**, 117 (2006a), [arXiv:astro-ph/0512082 \[astro-ph\]](#).
- X.-H. Fan, C. Carilli, and B. G. Keating, *Ann.Rev.Astron.Astrophys.* **44**, 415 (2006b), [arXiv:astro-ph/0602375 \[astro-ph\]](#).
- G. Hinshaw, D. Larson, E. Komatsu, D. Spergel, C. Bennett, *et al.*, (2012), [arXiv:1212.5226 \[astro-ph.CO\]](#).
- C. J. Hogan and M. J. Rees, *MNRAS* **188**, 791 (1979).
- D. Scott and M. J. Rees, *mnras* **247**, 510 (1990).
- M. McQuinn, O. Zahn, M. Zaldarriaga, L. Hernquist, and S. R. Furlanetto, *Astrophys.J.* **653**, 815 (2006), [arXiv:astro-ph/0512263 \[astro-ph\]](#).
- P. Madau, A. Meiksin, and M. J. Rees, *ApJ* **475**, 429 (1997), [arXiv:astro-ph/9608010](#).
- M. Zaldarriaga, S. R. Furlanetto, and L. Hernquist, *Astrophys.J.* **608**, 622 (2004), [arXiv:astro-ph/0311514 \[astro-ph\]](#).
- S. Furlanetto, M. Zaldarriaga, and L. Hernquist, *Astrophys.J.* **613**, 16 (2004a), [arXiv:astro-ph/0404112 \[astro-ph\]](#).
- S. Furlanetto, M. Zaldarriaga, and L. Hernquist, *Astrophys.J.* **613**, 1 (2004b), [arXiv:astro-ph/0403697 \[astro-ph\]](#).
- S. Furlanetto, S. P. Oh, and F. Briggs, *Phys.Rept.* **433**, 181 (2006), [arXiv:astro-ph/0608032 \[astro-ph\]](#).
- A. Lidz, O. Zahn, S. Furlanetto, M. McQuinn, L. Hernquist, *et al.*, *Astrophys.J.* **690**, 252 (2009), [arXiv:0806.1055 \[astro-ph\]](#).
- R. Wiersma, B. Ciardi, R. Thomas, G. Harker, S. Zaroubi, *et al.*, (2012), [arXiv:1209.5727 \[astro-ph.CO\]](#).
- M. A. Alvarez, E. Komatsu, O. Dore, and P. R. Shapiro, *Astrophys.J.* **647**, 840 (2006), [arXiv:astro-ph/0512010 \[astro-ph\]](#).
- R. Salvaterra, B. Ciardi, A. Ferrara, and C. Baccigalupi, *MNRAS* **360**, 1063 (2005), [arXiv:astro-ph/0502419](#).
- P. Adshead and S. Furlanetto, *Mon.Not.Roy.Astron.Soc.* (2007), [arXiv:0706.3220 \[astro-ph\]](#).
- H. Tashiro, N. Aghanim, M. Langer, M. Douspis, and S. Zaroubi, *MNRAS* **389**, 469 (2008), [arXiv:0802.3893](#).
- H. Tashiro, N. Aghanim, M. Langer, M. Douspis, S. Zaroubi, *et al.*, *Mon.Not.Roy.Astron.Soc.* **402**, 2617 (2010), [arXiv:0908.1632 \[astro-ph.CO\]](#).

- H. Tashiro, N. Aghanim, M. Langer, M. Douspis, S. Zaroubi, et al., (2011), arXiv:1008.4928 [astro-ph.CO].
- G. Holder, I. T. Iliev, and G. Mellema, *Astrophys.J.Lett.* (2006), arXiv:astro-ph/0609689 [astro-ph].
- V. Jelić, S. Zaroubi, N. Aghanim, M. Douspis, L. V. E. Koopmans, M. Langer, G. Mellema, H. Tashiro, and R. M. Thomas, *MNRAS* **402**, 2279 (2010), arXiv:0907.5179 [astro-ph.CO].
- A. Natarajan, N. Battaglia, H. Trac, U. L. Pen, and A. Loeb, (2012), arXiv:1211.2822 [astro-ph.CO].
- C. Dvorkin, W. Hu, and K. M. Smith, *Phys.Rev.* **D79**, 107302 (2009), arXiv:0902.4413 [astro-ph.CO].
- C. Dvorkin and K. M. Smith, *Phys.Rev.* **D79**, 043003 (2009), arXiv:0812.1566 [astro-ph].
- V. Gluscevic, M. Kamionkowski, and D. Hanson, (2012), arXiv:1210.5507 [astro-ph.CO].
- K. Masui, E. Switzer, N. Banavar, K. Bandura, C. Blake, et al., (2012), arXiv:1208.0331 [astro-ph.CO].
- E. Chapman, F. B. Abdalla, J. Bobin, J.-L. Starck, G. Harker, et al., (2012), arXiv:1209.4769 [astro-ph.CO].
- G. Harker, S. Zaroubi, G. Bernardi, M. A. Brentjens, A. de Bruyn, et al., *Mon.Not.Roy.Astron.Soc.* **405**, 2492 (2010), arXiv:1003.0965 [astro-ph.CO].
- G. Mellema, L. Koopmans, F. Abdalla, G. Bernardi, B. Ciardi, et al., (2012), arXiv:1210.0197 [astro-ph.CO].
- A. Liu, M. Tegmark, J. Bowman, J. Hewitt, and M. Zaldarriaga, (2009), arXiv:0903.4890 [astro-ph.CO].
- V. Jelić, S. Zaroubi, P. Labropoulos, R. M. Thomas, G. Bernardi, M. A. Brentjens, A. G. de Bruyn, B. Ciardi, G. Harker, L. V. E. Koopmans, V. N. Pandey, J. Schaye, and S. Yatawatta, *MNRAS* **389**, 1319 (2008), arXiv:0804.1130.
- G. Bernardi, A. B. G. Harker, M. Brentjens, B. Ciardi, V. Jelić, et al., (2010), arXiv:1002.4177 [astro-ph.CO].
- G. Bernardi, A. de Bruyn, M. Brentjens, B. Ciardi, G. Harker, et al., (2009), arXiv:0904.0404 [astro-ph.CO].
- A. Liu, J. R. Pritchard, M. Tegmark, and A. Loeb, (2012), arXiv:1211.3743 [astro-ph.CO].
- X. Wang and W. Hu, *Astrophys.J.* **643**, 585 (2006), arXiv:astro-ph/0511141 [astro-ph].
- G. B. Field, *Proceedings of the Institute of Radio Engineers.* **46**, 240 (1958).
- A. Lewis, A. Challinor, and A. Lasenby, *Astrophys.J.* **538**, 473 (2000), arXiv:astro-ph/9911177 [astro-ph].
- O. Zahn, A. Lidz, M. McQuinn, S. Dutta, L. Hernquist, et al., *Astrophys.J.* **654**, 12 (2006), arXiv:astro-ph/0604177 [astro-ph].
- M.-S. Shin, H. Trac, and R. Cen, *ApJ* **681**, 756 (2008), arXiv:0708.2425.
- R. K. Sheth and G. Tormen, *MNRAS* **329**, 61 (2002), arXiv:astro-ph/0105113.
- D. S. Reed, R. Bower, C. S. Frenk, A. Jenkins, and T. Theuns, *MNRAS* **374**, 2 (2007), arXiv:astro-ph/0607150.
- T. Baldauf, U. Seljak, R. E. Smith, N. Hamaus, and V. Desjacques, *ArXiv e-prints* (2013), arXiv:1305.2917 [astro-ph.CO].
- P. Shaver, R. Windhorst, P. Madau, and A. de Bruyn, (1999), arXiv:astro-ph/9901320 [astro-ph].
- A. Liu and M. Tegmark, *Mon.Not.Roy.Astron.Soc.* **419**, 3491 (2012), arXiv:1106.0007 [astro-ph.CO].
- A. Kogut, J. Dunkley, C. Bennett, O. Dore, B. Gold, et al., *Astrophys.J.* **665**, 355 (2007), arXiv:0704.3991 [astro-ph].
- L. La Porta, C. Burigana, W. Reich, and P. Reich, (2008), arXiv:0801.0547 [astro-ph].
- M. G. Santos, A. Cooray, and L. Knox, *Astrophys.J.* **625**, 575 (2005), arXiv:astro-ph/0408515 [astro-ph].
- X.-M. Wang, M. Tegmark, M. Santos, and L. Knox, *Astrophys.J.* **650**, 529 (2006), arXiv:astro-ph/0501081 [astro-ph].
- G. Harker, S. Zaroubi, G. Bernardi, M. A. Brentjens, A. G. de Bruyn, B. Ciardi, V. Jelić, L. V. E. Koopmans, P. Labropoulos, G. Mellema, A. Offringa, V. N. Pandey, J. Schaye, R. M. Thomas, and S. Yatawatta, *MNRAS* **397**, 1138 (2009), arXiv:0903.2760 [astro-ph.CO].
- E. Chapman, F. B. Abdalla, G. Harker, V. Jelić, P. Labropoulos, S. Zaroubi, M. A. Brentjens, A. G. de Bruyn, and L. V. E. Koopmans, *MNRAS* **423**, 2518 (2012), arXiv:1201.2190 [astro-ph.CO].
- E. Chapman, F. B. Abdalla, J. Bobin, J.-L. Starck, G. Harker, V. Jelić, P. Labropoulos, S. Zaroubi, M. A. Brentjens, A. G. de Bruyn, and L. V. E. Koopmans, *MNRAS* **429**, 165 (2013), arXiv:1209.4769 [astro-ph.CO].
- G. de Zotti, M. Massardi, M. Negrello, and J. Wall, *A&A Rev.* **18**, 1 (2010), arXiv:0908.1896 [astro-ph.CO].
- Planck Collaboration, P. A. R. Ade, N. Aghanim, F. Argüeso, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, and et al., *A&A* **536**, A13 (2011), arXiv:1101.2044 [astro-ph.CO].
- I. Heywood, M. J. Jarvis, and J. J. Condon, *MNRAS* **432**, 2625 (2013), arXiv:1302.2010 [astro-ph.CO].
- Planck Collaboration, P. A. R. Ade, N. Aghanim, M. I. R. Alves, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, and et al., *ArXiv e-prints* (2013), arXiv:1303.5073 [astro-ph.GA].
- M. F. Morales, *Astrophys.J.* **619**, 678 (2005), arXiv:astro-ph/0406662 [astro-ph].
- Y. Mao, M. Tegmark, M. McQuinn, M. Zaldarriaga, and O. Zahn, *Phys.Rev.* **D78**, 023529 (2008), arXiv:0802.1710 [astro-ph].
- M. P. van Haarlem et al., *A&A* **556**, A2 (2013), arXiv:1305.3550 [astro-ph.IM].
- (2006), arXiv:astro-ph/0604069 [astro-ph].
- M. Niemack, P. Ade, J. Aguirre, F. Barrientos, J. Beall, et al., *Proc.SPIE Int.Soc.Opt.Eng.* **7741**, 77411S (2010), arXiv:1006.5049 [astro-ph.IM].
- J. Austermann, K. Aird, J. Beall, D. Becker, A. Bender, et al., *Proc.SPIE Int.Soc.Opt.Eng.* **8452**, 84520E (2012), arXiv:1210.4970 [astro-ph.IM].
- M. Zaldarriaga, L. Colombo, E. Komatsu, A. Lidz, M. Mortonson, et al., (2008), arXiv:0811.3918 [astro-ph].
- M. Su, A. P. Yadav, M. McQuinn, J. Yoo, and M. Zaldarriaga, (2011), arXiv:1106.4313 [astro-ph.CO].
- H. V. Peiris and D. N. Spergel, *Astrophys.J.* **540**, 605 (2000), arXiv:astro-ph/0001393 [astro-ph].
- M. Tegmark, A. Taylor, and A. Heavens, *Astrophys.J.* **480**, 22 (1997), arXiv:astro-ph/9603021 [astro-ph].
- M. J. Mortonson and W. Hu, *Astrophys.J.* **657**, 1 (2007), arXiv:astro-ph/0607652 [astro-ph].
- O. Zahn, A. Mesinger, M. McQuinn, H. Trac, R. Cen, et al., (2010), arXiv:1003.3455 [astro-ph.CO].